DOCUMENT RESUME

ED 435 140 EA 030 124

AUTHOR Izumi, Lance T.
TITLE Developing and Implementing Academic Standards: A Template for Legislative and Policy Reform.
INSTITUTION Pacific Research Inst. for Public Policy, San Francisco, CA.
PUB DATE 1999-01-00
NOTE 74p.
AVAILABLE FROM Publications Department, Pacific Research Institute, 755 Sansome Street, #450, San Francisco, CA 94111 (paperbound: $16.95). Tel: 415-989-0833; E-mail: pripp@pacificresearch.org; Web site: <http://www.pacificresearch.org>.
PUB TYPE Guides - Non-Classroom (055)
EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Academic Standards; *Educational Change; Educational Improvement; Elementary Secondary Education; Evaluation; Performance Based Assessment; School Effectiveness; School Policy

ABSTRACT

This template is part of a series that recommends best practices and highlights problems that arise when policy deviates from these best practices. The booklet defines and outlines the critical components of a successful academic standards system using examples of good and bad standards to illustrate key points. The template addresses such questions as: What are the ingredients for rigorous academic standards? How do we measure compliance with standards? and How do performance standards differ from academic content standards? It offers analyses, discussion, and recommendations in four areas: tests aligned to the standards; categorization of student achievement through performance standards; implementation and accountability systems to guarantee that the standards become a reality in the classroom; and communications strategies to disseminate information about the standards to parents, teachers, and local school officials. The template is divided into four sections: (1) academic content standards; (2) the assessment system; (3) performance standards; and (4) implementation/accountability strategies. The object of the template is to offer policymakers practical advice and recommendations for constructing and implementing the best standards possible. It outlines the components of good academic content standards, such as rigor, intelligibility, and balance, and it reminds policymakers to consider such issues as depth versus breadth, time and cost of scoring, and standardization. Contains a summary fact sheet. (RJM)

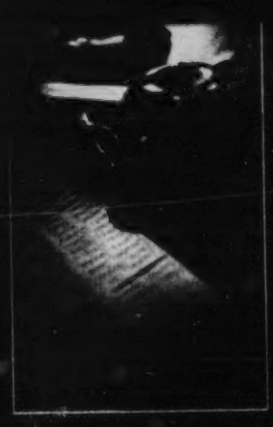
Reproductions supplied by EDRS are the best that can be made
from the original document.

PACIFIC RESEARCH INSTITUTE
1979 20 Years of Ideas 1999

ED 435 140

Developing and Implementing ACADEMIC STANDARDS

A Template for Legislative & Policy Reform



U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
L. Izumi
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
1

BA 030124

BEST COPY AVAILABLE

By Lance T. Izumi

2

Good Standards

Japan's 6th grade quantities and measurement standards include the following standards:

> Objective: To help children become able to measure the volume of fundamental solid figures. Furthermore, to help children know about the system of units of measuring and become able to efficiently measure the quantities. Content:
>
> 1) To enable children to measure the volume of fundamental solid figures through experiments and actual measurements, etc.
> a) To know how to measure the volume and surface area of fundamental prisms and circular cylinders.
> b) To know how to measure the volume of fundamental pyramids and circular cones. Furthermore, to know how to measure their surface area in simple cases.
> 2) To enable children to deepen their understanding of the measurements and units of quantities and to further develop their abilities to measure.
> a) To efficiently measure by using the proportional relationships.
> b) To understand the metric system and relations among their units and to efficiently use them in measurement.
>
> *(Japan; 6th grade quantities and measurement standards)*

The Japanese use a combination objective-content format for their math standards. The statement of objectives outlines clearly and specifically the learning goal, in this case, the measurement of volume and understanding the system of units of measurement. The objective is then followed with even more specific content requirements, in this case, measurement of the volume and surface areas of prisms, cylinders, pyramids, and cones, etc. After reading this set of standards, any teacher, parent, or member of the public would be confident of what is expected of the student in the classroom. There is no need for subtle interpretation or trying to figure out jargon-laden generalities. Even someone who didn't have the foggiest notion of what constitutes a "prism" would at least know that the student's task is to be able to measure its volume and surface area.

Developing and Implementing Academic Standards

A Template for Legislative and Policy Reform

By Lance T. Izumi*

Pacific Research Institute for Public Policy, San Francisco, California

January 1999

*Lance T. Izumi is co-director, Center for School Reform, Pacific Research Institute.

Copyright 1999 by the Pacific Research Institute for Public Policy
755 Sansome Street, Suite 450, San Francisco, CA 94111. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopy, recording, or otherwise, without prior written consent of the publisher.

Designed by Dawn Khalil

Table of Contents

About Pacific Research Institute and the Center for School Reform v

Acknowledgements vi

Introduction 1

Academic Content 3
- Rigor 4
- Intelligibility 8
- Measurability 10
- Specificity 12
- Comprehensiveness 15
- Academic 21
- Balance 24
- Manageability 26
- Cumulativeness 28

Assessment 31
- NAEP Assessment Development Process 31
- What Type of Assessment? 32
- Scoring 35
- Score Reporting 37
- Student Work Samples 38

Performance Standards 42

Implementation and Accountability 47

Communications 55

Summary 57

Endnotes 59

About the Author 62

About Pacific Research Institute and the Center for School Reform

The **Pacific Research Institute for Public Policy** promotes the principles of individual freedom and personal responsibility. The Institute believes these principles are best encouraged through policies that emphasize a free economy, private initiative, and limited government. By focusing on public policy issues such as education, the environment, law, economics, and social welfare, the Institute strives to foster a better understanding of the principles of a free society among leaders in government, academia, the media, and the business community.

The Institute's **Center for School Reform** explores and promotes a variety of school reform strategies, including charter schools, expanded school choice, public-private partnerships, school finance reform, raised academic standards, and well-crafted accountability systems.

Developing and Implementing Academic Standards is the first publication in the Center's series, **Templates for Legislative and Policy Reform**. The Center's Templates describe all relevant issues that officials and legislators must understand and incorporate in fashioning good policy in three key areas of education reform: performance and content standards, school choice, and charter schools. Drawing on experience from states that have dealt with these issues and from an exhaustive review of international and national data, the Templates detail how different methods of addressing each issue affect the outcome of specific legislation and policy. The Templates recommend best practices and language to ensure that desired outcomes are achieved and highlight problem areas that arise when policy deviates from those best practices. The three-part series includes:

- **Developing and Implementing Academic Standards**, by Lance T. Izumi
- **Making School Choice Work for All Families**, by John E. Coons and Stephen D. Sugarman
- **Expanding the Charter School Idea**, by Kenneth Lloyd Billingsley and Pamela Riley

Acknowledgements

The Pacific Research Institute and the Center for School Reform would like to thank the Gleason Works Foundation, the Oberndorf Family Fund, the Bernard Lee Schwartz Foundation, and the Walton Family Foundation for their very generous support. They underwrote the research and made it possible for Pacific Research Institute to publish the Templates for Legislative and Policy Reform.

Introduction

Unlike other education reform proposals, there is nearly universal agreement that education standards can be an important tool in improving student achievement. Standards inform students and their parents what society considers essential knowledge that children should learn during their K-12 education. Standards also provide taxpayers with benchmarks to judge how well the public schools are performing. For their part, lawmakers like standards because, at least initially, they involve relatively minor costs while affording elected officials the opportunity to claim that they are doing something concrete to improve the quality of public education.

Little wonder, then, that many states have so far adopted or are in the process of adopting some form of education standards that they expect their students to meet.

As with many promising ideas, however, the devil is truly in the details. The quality of standards so far adopted by states range from the rigorous and excellent to the vague and useless. It is a sad fact, unfortunately, that states have adopted more of the latter than of the former. Bad standards are worse than no standards at all because they cover up shortcomings in classroom instruction and student performance and, therefore, end up deceiving parents and the public.

Further, even where first-class standards have been approved, the mechanisms developed to implement those standards vary markedly in effectiveness. Also, bad assessment devices and poor performance standards can sabotage good content standards. Thus, for example, a rigorous set of standards can be undermined by an assessment device that deemphasizes the importance of students getting the right answer. Given such an assessment device, classroom teachers would have little incentive to teach to the standards.

With such significant potential for missteps in the entire standards process, the Pacific Research Institute for Public Policy has published this template on education standards that will give lawmakers, education officials, and the public a guide to the crafting of top-notch academic content standards, assessment devices, and performance standards, plus effective methods of implementing standards. This template addresses questions such as: How does one tell a good standard from a bad standard? What are the ingredients for rigorous academic standards? How

The goal of this template is to identify:

- **Good standards from bad.**
- **Ingredients for rigorous academic standards.**
- **Measurements for compliance.**
- **Performance standards vs. academic content.**
- **Best incentive systems in implementing standards.**

do we measure compliance with standards? How do performance standards differ from academic content standards? Which incentive system works best in implementing standards?

To provide answers to these and other key questions, the template is divided into the following sections:

- § Academic Content Standards
- § The Assessment System
- § Performance Standards
- § Implementation/Accountability Strategies

Again, the object of the template is not to rank standards since that has already been done quite well by the American Federation of Teachers and others, but rather to offer policymakers practical advice and recommendations for constructing and implementing the best standards possible.

Finally, it is important to point out that even good standards are not a panacea for all the ills of today's public education system. Standards can set challenging, yet reachable learning goals for children, but they cannot, by themselves, change key obstacles to real learning such as teacher training programs that fail to require subject-area competency.

Further, good standards cannot change structural problems such as collective bargaining processes that produce agreements that make it next to impossible to fire incompetent teachers.

Last, even standards that effectively raise the knowledge levels of students cannot guarantee that students will become more compassionate or moral beings. To achieve that would require adherence to a different set of standards authored by a higher authority than a legislature or board of education.

All that being said, however, there is much that a good standards system can accomplish. What follows, then, is an outline of what such a system should look like.

Academic Content

Unlike other education reforms such as charter schools and school choice (both of which are the subjects of separate PRI templates), one cannot lay out a single "model standard" as one can, for example, model school-choice legislation. Given that there are numerous subjects that students need to master, the impracticalities of such an effort are obvious. Instead, what can be produced is a framework around which all good standards should be built. A framework would be most useful for standards crafters.

During the 1990s, a consensus has developed on the qualities that should be embodied in any good set of standards. Citing past work by experts such as Paul Gagnon, Matt Gandal, and Chester Finn, standards analysts Dennis Doyle and Susan Pimental observe that a good set of standards, in whatever subject, should be:

1) Rigorous
2) Intelligible
3) Measurable
4) Specific
5) Comprehensive
6) Academic
7) Balanced
8) Manageable
9) Cumulative.[1]

These nine characteristics provide the needed framework for the construction of good standards. Before discussing each of these characteristics in detail, however, a short word about the documents will be used to illustrate and highlight the points made in each section. When policymakers first decide to put together sets of standards, they almost always turn their attention first to math and reading. This is not surprising since math and reading test scores are of great interest to parents, the public, elected officials, and the media. It is a happy coincidence that examples of standards in both subjects lend themselves very well to illustrating the strengths and shortcomings of various standards-crafting methods and philosophies.

Further, in both these two subjects, huge philosophical wars have been fought over how and what should be taught. In math, for example, factions battle over different methods of instruction. Clashes over phonics instruction versus "whole language" instruction have marked the reading debate. Thus, this paper will use several state math and reading standards to illustrate what works and what doesn't.

To date, a number of groups has issued rankings of state standards. In math, California, which just approved its math standards earlier this year (but only after intense battles between contending math forces), has received some of the highest marks from experts at the American Federation of Teachers and at think tanks such as the Fordham Foundation. Internationally, Japan's math standards (adopted in 1990 and officially designated as a curriculum) are rightfully viewed as some of the best in the world. As one set of reviewers observed, Japan's math standards "are refreshing to read; they are models of 'clear, definite, testable,' and no grade [standards] runs more than a few pages."[2] At the opposite end of the spectrum are the failed math standards of Michigan and Massachusetts. In reading, again, California's standards are widely viewed as among the best, while New Jersey's standards rank near the bottom. Also, reference will be made to Britain's National Curriculum (which is similar to standards in the United States). These nine standards, then, will serve as the illustrative tools for this part of the paper.

Rigor

Why are standards needed? The simple answer is that standards help increase student achievement by informing students as to what they need to know in order to progress through their schooling. For standards to increase student achievement, however, they must be demanding, i.e. they must set high expectations for students. Standards should ask students to reach their highest potential. In other words, they should be rigorous. Instead of being afraid to challenge students to excel for fear of bruising their self-esteem, standards should ask all children to master the core knowledge and skills necessary to compete with the best and brightest anywhere in the country or abroad.

Opponents of rigorous standards argue that it is unrealistic to expect many students, especially those from low-income families, to meet such standards. Yet, in cities and states where standards have been raised, students of all income and ethnic groupings have risen to the task. In New York City, for instance, after easier science classes were eliminated in 1994, the number of African-American ninth-graders passing the more difficult Regents science classes more than doubled by 1995, while the number of Latino ninth-graders passing Regents science quadrupled. In El Paso, Texas, where poverty is widespread, after academic requirements were raised in 1992 the number of Latino students taking and passing algebra courses increased significantly. In addition, the number of

Standards Characteristics
1) Rigorous
2) Intelligible
3) Measurable
4) Specific
5) Comprehensive
6) Academic
7) Balanced
8) Manageable
9) Cumulative

Latino students in El Paso passing state standardized math tests increased from 36.2% in 1992-93 to 86.4% in 1997-98.

The point is poor and minority students, like most other students, rise or fall to our level of expectation of them. All the more reason, therefore, to emphasize rigorous standards applicable to all students.

Bad Standards

The Massachusetts Pre-K to 4th grade number sense learning standards include the following standard:

> Students engage in problem solving, communicating, reasoning, and connecting to:
>
> 1) construct number meaning by using manipulatives and other physical materials to represent concepts of numbers in the real world,
> 2) demonstrate an understanding of our numeration system by relating counting, grouping, and place value concepts,
> 3) interpret the multiple uses of numbers by taking real-world situations and translating them into numerical statements. *(Massachusetts: Pre-K to 4th grade number sense learning standards)*

There are many reasons why this set of standards from Massachusetts fails the rigor test. First, instead of adopting grade-by-grade standards, Massachusetts has grouped six grades—pre-kindergarten, kindergarten, first, second, third, and fourth grades—and created a single set of standards that covers all these grades. The problem, of course, is that with such a large grouping of grades, the standards have to be very broad with little detail or specificity.

Take, for example, the second standard requiring students to demonstrate an understanding of concepts such as counting, grouping, and place value. Certainly students should be able to work with these concepts, but how much farther along should a student in the fourth grade be than a student in kindergarten or the first grade? If a kindergarten student should be able to count single and lower double-digit numbers, should a fourth grade student be able to count up into the thousands, ten-thousands, hundred-thousands, or millions?

By the fourth grade, should a student know decimal notation, and, if so, should it be to the tenths, hundredths, or thousandths? Should a fourth-grade student know what a negative number is? Given the overly broad language of the standard, it is impossible to answer any of these questions. (This overbroadness of language also afflicts the other two standards.) In addition, one notes that the entire set of Massachusetts

number sense standards for six grades consists of only three standards (an average of one standard for every two grades).

Compare that with the more than sixty-five detailed and specific number sense standards contained in California's math standards for grades K-4 (California uses a grade-by-grade system of standards). By any measure, then, there is a lack of rigor in the Massachusetts math standards that is due to, among other things, the decision to eschew grade-by-grade standards in favor of creating common standards to cover a large grouping of grades, the lack of detail and specificity, and the lack of comprehensiveness.

Good Standards

The California 4th grade number sense standards include the following standards:

1) Students understand place value of whole numbers and decimals to two decimal places, how these relate to simple fractions, and use concepts of negative numbers.
   - 1.1) read and write whole numbers in the millions,
   - 1.2) order and compare whole numbers and decimals to two decimal places,
   - 1.3) round whole numbers through the millions to the nearest ten, hundred, thousand, ten thousand, or hundred thousand,
   - 1.4) decide when a rounded solution is called for, and explain why this the case,
   - 1.5) interpret different meanings for fractions including parts of a whole, parts of a set, indicated division of whole numbers and quantities (and measures) between whole numbers on a number line; and relate to simple decimals on a number line,
   - 1.6) write tenths and hundredths in a decimal and fraction notation and know fraction/decimal equivalents for halves and fourths (e.g., 1/2 = 0.5 or .50; 7/4 = 1 3/4 = 1.75),
   - 1.7) write the fraction represented by a drawing of parts of a figure; represent a given fraction using drawings,
   - 1.8) use concepts of negative numbers (e.g., on a number line, in counting, in temperature, "owing"),
   - 1.9) identify the relative position of fractions, mixed numbers, and decimals to two decimal places on a line.
2) Students extend their use and understanding of whole numbers to addition and subtraction of simple decimals.
   - 2.1) estimate and compute the sum or difference of whole numbers and positive decimals to two places,
   - 2.2) round two place decimals to one decimal or the nearest whole number, and use rounding to judge the reasonableness of an answer.

Standards Characteristics
1) Rigorous
2) Intelligible
3) Measurable
4) Specific
5) Comprehensive
6) Academic
7) Balanced
8) Manageable
9) Cumulative

3) Students solve problems involving addition, subtraction, multiplication, and division of whole numbers, including the addition and subtraction of negative numbers, and understand the relationships among the operations.
   3.1) demonstrate understanding of, and the ability to use, standard algorithms for addition and subtraction of multi-digit numbers,
   3.2) demonstrate understanding of, and ability to use, standard algorithms for multiplying a multi-digit number by a two-digit number and long division for dividing a multi-digit number by a one digit number; use relationships between them to simplify computations and to check results,
   3.3) solve problems involving multiplication of multi-digit numbers by two-digit numbers and by one-digit numbers.
4) Students know how to factor small whole numbers.
   4.1) understand that many whole numbers decompose in different ways (e.g., 12 = 4 x 3 = 2 x 6 = 2 x 2 x 3),
   4.2) know that numbers such as 2, 3, 5, 7, and 11 do not have any factors except 1 and themselves, and that such numbers are called prime numbers.

(California; 4th grade number sense standard)

There is simply no comparison between this set of number sense standards for a single grade in California and the number sense standards for the six-grade grouping in Massachusetts. California's standards are complete—they give detailed information as to what students should know by the time they finish the fourth grade. Students must be able to count into the millions, know decimal places to the hundredths, write fractions and equate them to decimal notations, use negative numbers, compute sums and differences of numbers with decimals, do complex multiplication and division, etc. All of these requirements are built upon the equally complete and detailed number sense standards for each of the previous grades. Unlike in Massachusetts, students cannot slip by because of vague wording in the standards. The requirements are precisely worded and challenge students to understand and perform at a world-class level (indeed, the California standards are similar to Japan's, and in some cases, are even more demanding). In sum, the California standards are rigorous because they are complete, detailed, demanding, and cumulative. Further, California's decision to use grade-by-grade standards as opposed to grade groupings enhances each of these characteristics.

Intelligibility

Various authors make similar points about the need for intelligible standards, i.e. standards that are clear and understandable. Doyle and Pimental, for example, point out that one can tell whether standards are intelligible by the answers to questions such as:

> Are standards clear enough for teachers to understand what is required of them? Are standards written in jargon or are they clear enough for parents to keep an eye on their children's progress?[3]

A similar criterion is laid out by professors Ralph Raimi and Lawrence Braden who, in evaluating state math standards, define intelligibility/clarity as:

A. The words and sentences themselves must be understandable, syntactically unambiguous, and without needless jargon.
B. What the language says should be mathematically and pedagogically definite, leaving no doubt of what the inner and outer boundaries are, of what is being asked of the student or teacher.[4]

Bad Standards

The Michigan high school math standard for numerical and algebraic operations and analytical thinking includes the following standard:

> Explore problems that reflect the contemporary uses of mathematics in significant contexts and use the power of technology and algebraic and analytic reasoning to experience the ways mathematics is used in society.
> *(Michigan; high school math standard for numerical and algebraic operations and analytical thinking)*

Can a teacher, parent, or member of the public understand what this standard means? Most likely not. What does "reflect the contemporary uses of mathematics in significant contexts" mean? What is a "contemporary" versus a non-contemporary use of mathematics? How does one judge what are "significant contexts"? The recommendation of using "the power of technology and algebraic and analytic reasoning to experience the ways mathematics is used in society" sounds impressive, but it is unclear what it means, substantively and specifically. Given the questions in meaning raised by this standard, it is difficult to see how the standard could be used by educators and the public as a guide to what children should be learning.

Measurability

In order for standards to be effective in raising student achievement, they must be measurable, i.e. subject to being assessed. Later in this paper there will be a lengthy discussion of the issue of assessment. For assessment to work, assessment devices must be linked to the standards and the standards must be crafted so that they require students to produce work products capable of being measured. The standards, therefore, must eschew vagueness and emphasize the use of certain action verbs. Susan Munroe and Terry Smith, in their evaluation of state geography standards, list types of verbs that lend themselves to measurability:

> Standards employ strong verbs such as analyze, compare, demonstrate, describe, evaluate, explain, identify, illustrate, locate, make, trace, utilize, etc.[5]

As Doyle and Pimental note, verbs such as these "indicate an assessable action."[6]

The contrast between good assessable standards and useless non-assessable standards is perhaps starkest in mathematics. In math, there is almost always just one right answer and it is the pursuit of that right answer that is the goal of good standards, while bad standards de-emphasize getting the right answer lest students, supposedly, get discouraged.

Bad Standards

Michigan's elementary grade data analysis and statistical standards include the following standard:

> Collect and explore data through counting, measuring, and conducting surveys and experiments.
> *(Michigan; elementary grade data analysis and statistical standard)*

Contrast this standard with the first Japanese standard below. While the Japanese standard requires students to count and order objects correctly, the Michigan standard asks students to "collect and explore" data through counting, etc. What kind of data is to be collected? Data itself can be credible or flawed. More important, what does "explore" mean? How does one measure "explore"? Even if a child were to "collect and explore" data exactly as the standard writer envisioned, it would be impossible for an assessment device to measure whether the child had successfully accomplished the task because right and wrong answers do not flow from the standard. Children are merely given an ambiguous task which can be achieved by any type of performance.

Michigan officials would likely retort that they leave it up to the

local classroom teacher to give greater meaning to the standard. Local teachers would be expected to guide students in their data exploration. Yet, even if local teachers do engage in such an activity, the standard remains immeasurable. How could a state assessment device be created that sought to measure activities that might differ markedly from teacher to teacher? Aligning the assessment device to the state standards (which is critical if a standards system is to be effective) would, therefore, be impossible.

Good Standards

Japan's 1st grade numbers and calculation standards include the following standard:

> To correctly count or represent the number and order of objects. *(Japan; 1st grade numbers and calculation standard)*

Rather than using vague words to describe what children should be required to do, Japan's math standards repeatedly require students to do activities "correctly." In other words, they must not just count how many marbles are in a bag, they must count correctly and arrive at the right answer at the end of the activity. Requiring that the correct answer be the end product of the activity makes this standard easily measurable—either the student counts and orders objects and arrives at the correct answer or he does not. An assessment device can easily discover whether this standard is being met.

Japan's 4th grade numbers and calculations standards include the following standard:

> To know that the quotient is not changed if divisor and dividend are multiplied or divided by the same number as the property regarding division and use it in considering how to carry out the computation and checking the results of computing. *(Japan; 4th grade numbers and calculations standard)*

Japan also requires children to "know" things as facts. In this case, knowing that the answer to eight divided by two equals four is not changed by the eight and the two being both multiplied by, for instance, five. Once children know something as a fact or principle, then they are required to use that fact or principle in carrying out associated activities such as, in this standard, checking the results of computing. In terms of measurability, hard knowledge can be measured. One either knows or does not know that the quotient will be unaffected by multiplying or dividing the divisor and the dividend by the same number.

Standards Characteristics

1) Rigorous
2) Intelligible
3) Measurable
4) Specific
5) Comprehensive
6) Academic
7) Balanced
8) Manageable
9) Cumulative

Specificity

One of the main problems with many standards is that they are written in such a vague manner. Some defenders of vague state standards contend that the use of vague language gives local districts greater control over how and what is taught in the classroom. Of course, if that is the justification, why have a state standard in the first place? And if a state's education problem involves poorly run local public schools, then giving local districts, the very entities responsible for the low performance in the first place, the power to "fill in the gaps" will do little to raise student achievement. Indeed, in a penetrating critique of vague state standards, Chester Finn, Michael Petrilli, and Gregg Vanourek make the following insightful observation:

> The problem here is that vague standards are bound to serve as a barrier rather than a ladder to achievement. How can we expect students to master a body of knowledge if we fail to define what that body of knowledge is—and then convey it to them in a meaningful and accessible way? How can we monitor their progress toward benchmarks if we refuse to state those benchmarks in clear, identifiable, and measurable ways? How can we enlist the help of parents, volunteers, corporations, and others if nobody knows what they are supposed to be working towards? Vague standards set schools adrift without a map or compass—or even a destination.[7]

Standards, therefore, should be specific, while avoiding the peril of becoming mired in tiny details. Lawrence Lerner, in his evaluation of state science standards, gives a concise guide regarding the right level of specificity to which states should aim:

> [Standards] are specific but flexible; that is they are neither so broad as to be vague nor so narrow as to be trivial.[8]

Bad Standards

New Jersey's writing standards which students must achieve by the 4th grade include the following standards:

1) Use speaking, listening, reading, and viewing to assist with writing.
2) Write from experiences, thoughts, and feelings.
3) Use writing to extend experience.
4) Write for a variety of purposes, such as to persuade, enjoy, entertain, learn, inform, record, respond to reading, and solve problems.

5) Write on self-selected topics in a variety of literary forms.
6) Write collaboratively and independently.
7) Use a variety of strategies such as brainstorming, listening, discussion, drawing, role playing, note-taking, and journal writing, for finding and developing ideas about which to write. *(New Jersey: writing standards students must achieve by the 4th grade)*

These standards are so broad as to be virtually meaningless. For example, everyone uses speaking, listening, reading, and viewing in some way to help them with writing. Without some more specific requirement, this standard is about as helpful as telling someone to open one's eyes in order to see. The standards also ask students to write from experiences, thoughts, and feelings. But write what? And for what purpose? The standards ask students to "write for a variety of purposes, such as to persuade, enjoy, entertain, learn, inform, record, respond to reading, and solve problems." Well, that just about sums up almost every form of writing. This standard could have been shortened simply to: "Just go out there and write." Likewise, the standard, "Write on self-selected topics in a variety of literary forms," might as well say, "Write about anything you like any way you like." "Write independently or collaboratively"? That certainly exhausts all possibilities, i.e. one can only write by one's self or with another. The bottom line is that these standards are so general that they offer no real guidance for teachers or students. In addition, their excessive breadth makes them totally immeasurable.

New Jersey officials would likely argue that one of the reasons for the lack of specificity of these standards is to refrain from stifling the creativity of students. According to this point of view, students should be given the freedom to explore and experiment with many different ways of writing, and should be encouraged to do so by standards that avoid artificial straitjackets. The trouble with this argument, however, is that in order to communicate effectively through writing, students must be able to understand and use the structure of good written communication. Understanding and using such structure does not impede creativity, but rather gives students a means of organizing their thoughts so that their creativity can be enhanced. Ignoring the importance of such structure often leaves students frustrated and confused (states of being certainly not conducive to creativity).

Good Standards

California's 4th grade writing standards on organization and focus include the following standards:

1.1) Select a focus, an organizational structure, and a point of view based upon purpose, audience, length, and format requirements.
1.2) Create multiple-paragraph compositions:
  a) Provide an introductory paragraph.
  b) Establish and support a central idea with a topic sentence at or near the beginning of the first paragraph.
  c) Include supporting paragraphs with simple facts, details, and explanations.
  d) Conclude with a paragraph that summarizes the points.
  e) Use correct indention.
1.3) Use traditional structures for conveying information (e.g., chronological order, cause and effect, similarity and difference, and posing and answering a question).

(California; 4th grade writing standards on organization and focus)

In contrast to New Jersey, California's writing standards are models of specificity. For example, California requires that students write compositions based upon a recognized structure. Each part of that structure has a purpose that the student must meet (e.g., "Establish and support a central idea with a topic sentence at or near the beginning of the first paragraph.") and which others can use to measure whether students have or have not met the standard. Students cannot turn in any piece of paper with words scrawled on it and meet the standard (as is the case with New Jersey's standards). Rather, the specificity of the standards makes them measurable, and, therefore, valuable.

Also, California's 4th grade standards for written and oral language conventions include the following standards:

1.1) Use simple and compound sentences in writing and speaking.
1.2) Combine short, related sentences with appositives, participial phrases, adjectives, adverbs, and prepositional phrases.
1.3) Identify and use regular and irregular adverbs, prepositions, and coordinating conjunctions in writing and speaking.
1.4) Use parentheses, commas in direct quotations, and apostrophes in the possessive case of nouns and in contractions.
1.5) Use underlining, quotation marks, or italics to identify titles of documents.

1.6) Capitalize names of magazines, newspapers, works of art, musical compositions, organizations, and the first word in quotations when appropriate.
1.7) Spell correctly roots, inflections, suffixes and prefixes, and syllable constructions.
(California; 4th grade standards for written and oral English language conventions)

Again, like the previous example, these California standards are very specific. They require students to know various word types and their use, punctuation and its uses, capitalization and its uses, etc. Because of this specificity, any good assessment device would be able to measure whether students had met these standards or not.

Comprehensiveness

Effective standards must adequately cover all essential and important areas of the subject. Doyle and Pimental ask:

> Are the standards complete? Do they cover the subject in adequate breadth and depth? Do standards contain the major concepts of a field, the essential ideas that students must master if they are to have a grasp of the field?[9]

Consider a set of standards that left gaps in its first grade standards but was comprehensive in its second grade standards. The result would be disastrous. E.D. Hirsch, editor of the Core Knowledge book series, emphasizes the importance of shared knowledge among all students and how that shared knowledge correlates to student performance:

> If shared knowledge is needed among citizens to understand newspapers as well as one another, then, by the same reasoning, shared knowledge is also needed among class members to understand the teacher and one another. Every classroom is a little society of its own, and its effectiveness and fairness depend on the full participation by all its members, just as in the larger society. Such universal participation by students cannot occur unless they all share a core of relevant background knowledge. This is easily demonstrated in any classroom group. To the extent that lack of relevant knowledge keeps some students from comprehending today's lesson, it will cause them to fall even further behind in comprehending tomorrow's.[10]

Standards with gaps in their knowledge could, therefore, end up preventing legions of students from acquiring the relevant background knowledge they will need to handle harder material in future grades.

Bad Standards

New Jersey's reading standards which students must achieve by the 4th grade include the following standards:

1) Use listening, speaking, writing, and viewing to assist with reading.
2) Listen and respond to whole texts.
3) Understand that authors write for different purposes, such as persuading, informing, entertaining, and instructing.
4) Use reading for different purposes, such as enjoyment, learning, and problem-solving.
5) Read independently a variety of literature written by authors of different cultures, ethnicities, genders, and ages.
6) Read literally, inferentially, and critically.
7) Use print concepts in developmentally appropriate ways.
8) Read with comprehension.
9) Use prior knowledge to extend reading ability and comprehension and to link aspects of the text with experiences and people in their own lives.
10) Identify passages in the text that support their point of view.
11) Distinguish personal opinions and points of view from those of the author, and distinguish fact from opinion.
12) Demonstrate comprehension through retelling or summarizing ideas and following written directions.
13) Identify elements of a story, such as characters, setting, and sequence of events.
14) Identify literary forms, such as fiction, poetry, drama, and nonfiction.
15) Expand vocabulary using appropriate strategies and techniques, such as word analysis and context clues.
16) Read and use printed materials and technical manuals from other disciplines, such as science, social studies, mathematics, and applied technology.

(New Jersey; reading standards students must achieve by the 4th grade)

Standards Characteristics
1) Rigorous
2) Intelligible
3) Measurable
4) Specific
5) Comprehensive
6) Academic
7) Balanced
8) Manageable
9) Cumulative

These sixteen standards lack any standard dealing with phonics or how children best learn to read. The evidence is overwhelming that phonics works and must, therefore, be a major element in teaching children to read.[11] Yet, not a single standard anywhere deals with phonics or, for that matter, reading ability acquisition. New Jersey officials evidently assume competent reading ability on the part of all students in grades K-4 since students are asked to "Read with comprehension" and "Read literally, inferentially, and critically." One cannot read with comprehension without being able to read in the first place. Yet, in the "Descriptive Statement" preceding the reading standards, New Jersey sidesteps this crucial first step by saying that, "Proficient readers use a repertoire of strategies (including phonics, context clues, and foreshadowing) that enables them to adapt to increasing levels of complexity, and they develop lifelong habits of reading and thinking." In other words, their assumption seems to be that all reading learning methods are equally effective and result in reading proficiency that increases through the years. That view is obviously untrue, and New Jersey's failure to deal with the question of how children best acquire reading ability leaves a massive hole in its reading standards.

New Jersey's reading standards which students must achieve by the 12th grade include the following standards:

> 26) Understand the relationship between contemporary writing and past literary traditions.
> 27) Understand that our literary heritage is marked by distinct literary movements and is part of a global literary tradition.
> 28) Analyze how the works of a given period reflect historical events and social conditions.
> 29) Understand the study of literature and theories of literary criticism.
> 30) Understand appropriate literary concepts, such as rhetorical device, logical fallacy, and jargon.
> 31) Understand the effect of such literary devices, such as alliteration and figurative language, on the reader's emotions and interpretation.
> 32) Understand the range of literary forms and content that elicit aesthetic response.
> *(New Jersey; reading standards students must achieve by the 12th grade)*

By the time students graduate from high school, students should have read Ernest Hemingway, Nathaniel Hawthorne, James Fenimore Cooper, John Steinbeck, and other American literary giants. The closest that the New Jersey standards come to mentioning American literature is the allusion to "our literary heritage" in standard #27. Yet, even here,

the standard does not require students to read or analyze any works in the canon of American literature, but merely asks them to understand that "our literary heritage" is connected to a "global literary tradition." No need to read *The Grapes of Wrath* or any other book to meet that standard.

New Jersey officials would likely counter that they are simply giving flexibility to local schools and teachers to assign books that are "relevant" to students. For example, teachers in schools with large Hispanic student populations are free to assign books written by Central and South American authors. The point, though, is not that there is anything inherently wrong with students reading the works of such authors (there is not), rather the problem is that by the time students graduate from high school they should have read a significant number of works in the American literary canon. This does not mean that state standards should dictate a specific book list (they should not), but that a key area of knowledge such as American literature should be an explicit part of any reading standards, with flexibility given to local teachers as to which specific works they assign their pupils.

Good Standards

California's 1st grade reading standards on word analysis, fluency, and systematic vocabulary development include the following standards:

1.1) Match oral words to printed words.
1.2) Identify the title and author of a reading selection.
1.3) Identify letters, words, and sentences.
1.4) Distinguish initial, medial, and final sounds in single-syllable words.
1.5) Distinguish long- and short-vowel sounds in orally stated single-syllable words (e.g., bit/bite).
1.6) Create and state a series of rhyming words, including consonant blends.
1.7) Add, delete, or change target sounds to change words (e.g., change cow to how; pan to an).
1.8) Blend two to four phonemes into recognizable words (e.g., /c/a/t/ = cat; /f/l/a/t/ = flat).
1.9) Segment single syllable words into their components (e.g., /c/a/t/ = cat; /s/p/l/a/t/ = splat; /r/i/c/h/ = rich).
1.10) Generate sounds from all the letters and letter patterns, including consonant blends and long- and short-vowel patterns (i.e., phonograms), and blend those sounds into recognizable words.
1.11) Read common, irregular sight words (e.g., the, have, said, come, give, of).
1.12) Use knowledge of vowel digraphs and r-controlled

Standards Characteristics
1) Rigorous
2) Intelligible
3) Measurable
4) Specific
5) Comprehensive
6) Academic
7) Balanced
8) Manageable
9) Cumulative

letter-sound associations to read words.
1.13) Read compound words and contractions.
1.14) Read inflectional forms (e.g., -s, -ed, -ing) and root words (e.g., look, looked, looking).
1.15) Read common word families (e.g., -ite, -ate).
1.16) Read aloud with fluency in a manner that sounds like natural speech.
1.17) Classify grade-appropriate categories of words (e.g., concrete collections of animals, foods, toys).
(California; 1st grade reading standards on word analysis, fluency, and systematic vocabulary development)

Compare these seventeen standards with the sixteen standards that New Jersey expects its students to meet by the fourth grade. The California standards, for a single grade, are very detailed and specific. They do not assume students will pick up essential knowledge to fill in the gaps in the standards, an assumption made by New Jersey. Rather, the California standards offer a comprehensive set of assessable knowledge-based requirements that will give students the necessary knowledge tools they will need to read. The California standards cover critical areas such as concepts about print, phonemic awareness, decoding and word recognition, and vocabulary and concept development. The standards are also part of a comprehensive set of first-grade reading standards that include similarly detailed standards on reading comprehension and literary response and analysis. Further, subsequent standards in higher grades build on the knowledge students are asked to acquire in the first grade. The result is a set of standards that is complete, covering the subject in admirable depth and breadth.

California's 11th and 12th grade reading standards on literary response and analysis include the following standards:

3.1) Analyze characteristics of subgenres (e.g., satire, parody, allegory, pastoral) that are used in poetry.
3.2) Analyze the way in which the theme or meaning of a selection represents a view or comment on life, using textual evidence to support the claim.
3.3) Analyze the ways in which irony, tone, mood, the author's style, and the "sound" of language achieves specific rhetorical or aesthetic purposes, or both.
3.4) Analyze ways in which poets use imagery, personification, figures of speech, and sounds to evoke readers' emotions.
3.5) Analyze recognized works of American literature representing a variety of genres and traditions:
a) Trace the development of American literature from the colonial period forward.

b) Contrast the major periods, themes, styles, and trends and describe how works by members of different cultures relate to one another in each period.
c) Evaluate the philosophical, political, religious, ethical, and social influences of the historical period that shaped the characters, plots, and settings.

3.6) Analyze the way in which authors through the centuries have used archetypes drawn from myth and tradition in literature, film, political speeches, and religious writings (e.g., how the archetypes of banishment from an ideal world may be used to interpret Shakespeare's tragedy *Macbeth*).

3.7) Analyze recognized works of world literature from a variety of authors:
a) Contrast the major literary forms, techniques, and characteristics of the major literary periods (e.g., Homeric Greece, medieval, romantic, neoclassic, modern).
b) Relate literary works and authors to the major themes and issues of their eras.
c) Evaluate the philosophical, political, religious, ethical, and social influences of the historical period that shaped the characters, plots, and settings.

3.8) Analyze the clarity and consistency of political assumptions in a selection of literary works or essays on a topic (e.g., suffrage, women's role in organized labor). (Political approach)

3.9) Analyze the philosophical arguments presented in literary works to determine whether the authors' positions have contributed to the quality of each work and the credibility of the characters. (Philosophical approach)

(California; 11th and 12th grade reading standards on literary response and analysis)

As with California's first grade reading standards, the state's eleventh- and twelfth-grade reading standards are models of detail, specificity, and completeness. Unlike the New Jersey standards, the California standards require students to read and analyze "recognized works of American literature" from the colonial period to the present. Further, after reading these works, students must be able to recognize and display their knowledge of such key literary elements as themes and styles, plus relate the works to the various influences that marked the particular historical period (note the difference in specificity between New Jersey reading stan-

dard #28 and California's standard #3.5c). In other words, students must read and understand these important works. Notice, however, that while the California standards include American literature requirements, they do not dictate a list of which literary works must be read. This is positive since classroom teachers should be given some flexibility in making selections within the canon.

The California standards also require students to gain specific knowledge in areas such as the structural features of literature (3.1), narrative analysis of grade-level-appropriate text (3.2-3.7), and literary criticism (3.8-3.9). This entire section is also part of a larger set of reading standards for grades eleven and twelve which includes similarly detailed standards on word analysis, fluency, and systematic vocabulary development and reading comprehension. If students adhere to California's comprehensive standards, they will leave high school as literate members of society. If students in New Jersey leave high school in the same condition, it will be in spite of, not because of, that state's reading standards.

Academic

Although standards are supposed to be an educational tool, they are a tempting target for some who view them as vehicles to change students' social or political views or to modify behavior. It is, therefore, crucial that standards always remain focused on all-important academic concerns, and not end up becoming the Christmas tree upon which every ideologue, social engineer, or behaviorist hangs his or her favorite ornament.

Sandra Stotsky, in her evaluation of state English language-arts/reading standards, lists a number of anti-academic expectations to beware:

> The reading/literature standards require students to relate what they read to their lived experiences. The reading/literature standards want reading materials to address contemporary social issues. The document implies that all literary and nonliterary texts are susceptible of an infinite number of interpretations and that all points of view or interpretations are equally valid regardless of logic, accuracy, and adequacy of supporting evidence. The examples of classroom activities or student writing offered are politically slanted or reflect an attempt to manipulate students feelings, thinking, or behavior. The standards teach moral or social dogma.[12]

Bad Standards

Massachusetts' Guiding Principle II of Mathematics Education includes the following:

> All students have access to high quality mathematics program.... To help learners in Massachusetts schools reach their full potential, the diversity in communities and classrooms should be treated as an advantage. The presence of diverse learners in Massachusetts classrooms presents an opportunity for all students and teachers to learn about the rest of the world and appreciate the talents and culture of each individual. Since different cultures sometimes use alternative mathematical strategies or perceive the relationships of objects and events in the world in ways other than the mainstream culture, their strategies and understandings can enrich the understanding of all students. For example, Cambodian children learn a different algorithm for division. If given the opportunity to explain their method to the rest of the class, then everyone broadens their cultural experiences, deepens their understanding of the concept of division, and recognizes the varied approaches to mathematics.
> *(Massachusetts; Guiding Principle II of Mathematics Education)*

Massachusetts includes seven "guiding principles" of math education as a preface to its standards. Although not a standard, the guiding principles serve as "the underlying beliefs and tenets central to the vision of mathematical power and content standards for mathematics education in Massachusetts." The problem with this particular principle and its elaboration is that it is a variation on the typical bad standard which, Stotsky warns, accepts an infinite number of interpretations as equally valid "regardless of logic, accuracy, and adequacy of supporting evidence." In this case it is true that ethnic diversity in the classroom can be a very positive thing. But, that does not mean that the ways in which each culture views math should necessarily be given equal weight. Some ways of viewing math are certainly more successful than others. Students should learn these successful methods first, and then use them as benchmarks to evaluate other methods. It is questionable, however, whether students in Massachusetts will get such instruction since the Massachusetts math standards also say that schools need to move away from the notion that students should master the basics before proceeding to higher level mathematics.

Contrary to the seeming requirements of politically correct multiculturalism, students should be taught using the best methods available and not be confused into thinking that there is nothing wrong in using less successful methods. The ultimate goal of math standards and math education is not to form a better understanding of other cultures, better taught in cultural anthropology or social studies classes, but to impart math knowledge to students in the most effective way possible. The math classroom is no place for cultural and instructional relativism.

Good Standards

Japan's 3rd grade objective for arithmetic in elementary school math standards includes the following:

> 1) To help children become able to use decimal fractions and common fractions to represent the size of quantities. Furthermore, to help them understand the meanings of multiplication and division of whole numbers and become able to compute in basic calculations, as well as to help children appreciate their usefulness and become able to apply them exactly and efficiently according to their purposes.
> 2) To help children understand the concepts of weight and time and become able to measure the fundamental quantities such as length through appropriately choosing units and tools according to their purpose.
> 3) To help children deepen their understanding of fundamental geometrical figures and become able to construct and use them.
> 4) To help children become able to arrange data, and use mathematical expressions and graphs, and to help children appreciate their meaning and become gradually able to represent or investigate the sizes of quantities and their mathematical relations.
>
> *(Japan; Objective for 3rd Grade Arithmetic in elementary school math standards)*

In addition, Japan's objective for Mathematics I in math standards for upper secondary school includes the following:

> Through consideration of concrete phenomena, to help students understand quadratic functions, geometrical figures and mensuration, treatment of numbers of cases and probability, and to encourage them to master basic knowledge and skills, to develop their abilities to utilize them exactly, and to deepen their appreciation of the significance of the mathematical way of viewing and thinking.
> *(Japan; Objective for Mathematics I in math standards for upper secondary school)*

As mentioned earlier under the "Intelligibility" section, rather than guiding principles, Japan has elected to place a statement of objectives in front of the content standards for each of the primary- and middle-school grades, and in front of the content standards by math subject area for the high-school grades. As can be seen, these statements of objectives

Standards Characteristics
1) Rigorous
2) Intelligible
3) Measurable
4) Specific
5) Comprehensive
6) Academic
7) Balanced
8) Manageable
9) Cumulative

are academic statements which concentrate on content matter. There is no political correctness or multicultural relativism here and no confusion about what is expected from students. They are required to "compute in basic calculations," "measure fundamental quantities such as length," "arrange data," "master basic knowledge and skills," and to apply what they learn "exactly and efficiently." Further, as Raimi and Braden point out, Japan's content standards, which follow the statement of objectives, are specific enough that students will be inculcated with a "mathematical way of viewing and thinking."[13]

Balance

There is an important distinction between skill and knowledge. Skills involve the ability of individuals to perform tasks. Thus, being able to find books in the library using the library's filing code system is a skill. Knowledge, on the other hand, involves being able to comprehend and retain the content of the book in the library. While standards should not ignore skills, it is knowledge that should be emphasized. As Doyle and Pimental point out:

> How is it possible to decide, analyze, investigate, compare, or classify without content? Skills can't be taught in the abstract. Neither can they be assessed. Knowledge is the scaffolding upon which critical thinking is built.[14]

Yet, despite the importance of ensuring that students acquire good content knowledge, many standards give short shrift to knowledge requirements. Why? According to Finn, Petrilli, and Vanourek, the reason involves the belief in many elite education circles that there is no such thing as real knowledge:

> There is no actual truth or definite knowledge, relativists believe, only various culturally determined "scripts" or "versions" of the truth. It would be oppressive, they argue, for a state to identify specific knowledge that must be learned by all. Any such knowledge would be nothing more than script preferred by the dominant class. Better to leave it out altogether.[15]

Such elitist opinions notwithstanding, knowledge of essentials should form the central basis of any set of good standards. Stotsky, for instance, places great emphasis on core content knowledge in her English/reading standards evaluation:

> [The standards] include knowledge of diverse literary elements and genres, different kinds of literary responses,

> and use of a variety of interpretive and critical lenses. They also specify those authors, works, and literary traditions in American literature and in the literary and civic heritage of English-speaking people that all students should study because of their literary quality and cultural significance.[16]

Bad Standards

New Jersey's speaking standards that students must meet by the 4th grade include the following standards:

> 5) Participate in collaborative speaking activities, such as choral reading, plays, and reciting poems.
> 6) Participate in discussion by alternating the roles of speaker and listener.
> *(New Jersey: speaking standards that students must meet by the 4th grade)*

These standards ask students to participate in various activities, but they do not ask students to gain any real knowledge. For example, a student may recite a poem with a number of his classmates, but what is he or she supposed to get out of the recitation? What is he or she supposed to bring to the recitation to make it an activity of some intellectual value? While it may be useful to have students memorize a poem and then recite it back in class (many experts say memorization is beneficial), the New Jersey standards make no mention of memorization (a student could meet the standard by simply reciting the poem out of a book). Instead of asking students to engage in an activity intended to spur knowledge acquisition, these New Jersey standards merely ask students to hone a skill, e.g., being able to read a part of a play or poem orally.

Also, New Jersey's K-4 cumulative language arts and literacy standard on visual information includes the following:

> Take notes on visual information from films, presentations, observations, and other visual media, and report that information through speaking, writing, or their own visual representation.
> *(New Jersey: K-4 cumulative language arts and literacy standard on visual information)*

There is no question that note-taking is an important activity, but this standard provides no guidance regarding the knowledge that students should gain from the visual event they are observing. Should students be able to glean and distill the main ideas emanating from the visual event? According to this standard, a student could take down and

relate the descriptive information from the visual event, miss the event's main idea completely, and still satisfy the standard. For example, a student could watch a film of a Soviet May Day parade, diligently describe the soldiers and military hardware passing before the reviewing stands, and not address the parade's key objective of intimidating friends and foes alike. Note-taking without understanding, therefore, does not enhance the acquisition of knowledge.

Good Standards

California's 4th grade listening and speaking standards on organization and delivery of oral communication include the following standards:

> 1.5) Present effective introductions and conclusions that guide and inform the listener's understanding of important ideas and evidence....
> 1.7) Emphasize points in ways that help the listener or viewer to follow important ideas and concepts.
> 1.8) Use details, examples, anecdotes, or experiences to explain or clarify information.
> *(California; 4th grade listening and speaking standard on organzation and delivery of oral communication)*

Unlike the New Jersey standards, these California standards connect the skill activity to the acquisition or enhancement of knowledge. For example, being able to use introductions, conclusions, and emphasis in an oral presentation is connected to the knowledge component of pinpointing and following important ideas. Students must be able to know what the important ideas and concepts are in the first place, and the foundation of knowledge underlies the skill component. California's standards, therefore, call for a balance between the honing of a skill and the acquisition of knowledge that gives meaning to the skill.

Manageability

In their evaluation of geography standards, Munroe and Smith advise that good standards "offer guidance to teachers in developing curriculum activities, classroom materials, and instructional methods."[17] The key word here is "guidance." What standards should do is offer a road map for teachers to follow so they can figure out what essential core content knowledge they must convey and impart to their students. Outside of this core, teachers should be given some latitude in bringing in other information. In his evaluation of science standards Lawrence Lerner notes:

Standards Characteristics
1) Rigorous
2) Intelligible
3) Measurable
4) Specific
5) Comprehensive
6) Academic
7) Balanced
8) Manageable
9) Cumulative

> [Standards] comprehensively cover basic knowledge, the importance of which is generally agreed upon by the scientific community; they are not, however, encyclopedic.[18]

In other words, the size and scope of the standards should be manageable. Details that are too small or narrow can be placed in curriculum frameworks, teacher guides, or other documents.

Bad Standards

Ever since Britain adopted its National Curriculum at the beginning of the decade, British teachers have been plagued with manageability problems. The curriculum covers ten subjects (English, mathematics, science, technology, history, geography, art, music, foreign language, and physical education) in such detail that, according to one British researcher, "any desirable activity became compulsory."[19] The result has been that teachers are literally unable to cram all the requirements into the allotted teaching time. As a consequence, according to education professor James Tooley, head of the education unit at the London-based Institute for Economic Affairs, many British teachers, convinced of the need for more time devoted to reading and basic numeracy, are "coerced by experts instead to cram their curriculum with 'Blue Peter' technology and meandering investigations."[20] One British head teacher further observed:

> As everybody knows, each subject was developed separately. Given this approach we should not be surprised that those who were involved tried to grab as much human knowledge as they could and crammed it into their subject. This was inevitable. Taken subject by subject, this was bad enough but then they all also started to recommend what percentage of the total timetable should be devoted to their subject and the total came to 130 percent of the time available. This is nonsense. Somebody must take an overview. We need a framework but one in which there is some flexibility.[21]

No wonder that organizations such as the secondary headteacher's association have called for a slimming down of both the overall National Curriculum and each subject curriculum.[22] Also, last year, British chief inspector of schools Chris Woodhead called for a slim-down of the National Curriculum so that teachers could spend more time on literacy and numeracy.[23] The current paradox, though, remains that because the British National Curriculum tries to cover so much, students end up learning less.

Standards Characteristics

1) Rigorous
2) Intelligible
3) Measurable
4) Specific
5) Comprehensive
6) Academic
7) Balanced
8) Manageable
9) Cumulative

Good Standards

Though often cited for excellence, Japanese math standards are only forty-seven pages long. Yet those forty-seven pages contain the objectives and content standards for every grade up to twelve, honors tracks, a track for those interested in science or engineering, plus guidelines for the construction of teaching plans. As Raimi and Braden approvingly point out:

> All this is achieved in a mere forty-seven pages, less than a tenth of the longest American framework, though "framework" is probably a good description of the Japanese document as well, for it does contain pedagogical hints of importance.[24]

By keeping to the essentials and by not getting mired down in encyclopedic detail, the Japanese standards remain manageable for both teachers and students.

Cumulativeness

Learning is a building process. Knowledge gained in early grades forms the foundation for higher learning in succeeding grades. Standards writers must, therefore, ensure that knowledge requirements for lower grades are complete enough so that students will not be confused when faced with more difficult material. Conversely, the standards for higher grades must be challenging so as to take full advantage of the foundation of knowledge laid down in the lower grades.

Lawrence Lerner used as a key indicator of science standards quality the following criteria:

> [The standards] expect increasing intellectual sophistication and higher levels of abstraction, as well as the skills required to deal with increasingly complex arrays of information, at successively higher educational levels. In light of the tight logical structure of the sciences, it is especially important that the standards also expect knowledge gained by students to be cumulative, each level building on what has been mastered earlier.[25]

It is not just the "logical" sciences, though, that require this cumulative effect. Sandra Stotsky says this about English/reading standards:

> [The standards] are of increasing intellectual difficulty at each higher level and cover all important indices of learning in the area they address.[26]

Bad Standards

Michigan's elementary, middle-school, and high-school grade data analysis and statistics standards include the following standards:

- § Formulate questions and problems and gather and interpret data to answer those questions. *(Michigan; elementary grade data analysis and statistics standard)*
- § Formulate questions and problems and gather and interpret data to answer those questions. *(Michigan; middle-school grade data analysis and statistics standard)*
- § Formulate questions and problems and gather and interpret data to answer those questions. *(Michigan; high-school grade data analysis and statistics standard)*

No, this is not a misprint. For each of the three grade groupings, Michigan merely regurgitates the same vague standard. Even if the standard were used for just one of the grade groupings, it would be a travesty since this vague standard lacks rigor and is immeasurable and unspecific. To have the same vague standard for three grade groupings, however, creates the added problem that a child could conceivably formulate the same questions in the 12th grade as he or she did in the 1st grade and still literally meet the standards.

Michigan officials would no doubt say such is not their intent. They would argue that they do expect progression of knowledge, but that such progression should be demonstrated not by the wording of the state standard, but by samples of student work, expectations of classroom teachers, guidelines issued by local districts, etc. The trouble with this argument is that the rationale for state standards is not to create a void which must be filled by local teachers, principals, and district officials, but rather to create a set of concrete guideposts that inform teachers and others as to what children should be learning and achieving. Standards should not be a means of passing the buck.

Good Standards

Japan's numbers and calculations objectives for the 3rd through the 6th grades include the following:

- § To help children become able to use decimal fractions and common fractions to represent the size of quantities. Furthermore, to help them understand the meanings of multiplication and division of whole

numbers and become able to compute in basic calculations ....
(Japan; 3rd grade numbers and calculations objective)

§ To help children deepen their understanding of whole numbers and how to express decimal fractions and common fractions as well as understanding the meaning of round numbers .... Furthermore, to help them become able to master the four basic operations [addition, subtraction, multiplication, and division] with whole numbers and effectively apply to consideration of phenomena and use addition and subtraction of decimals and fractions.
(Japan; 4th grade numbers and calculations objective)

§ To help children understand the meaning of multiplication and division of decimal fractions and become able to compute in decimals and fractions, as well as become able to make use of them in considering phenomena....
(Japan; 5th grade numbers and calculations objective)

§ To help children understand the meaning of multiplication and division of fractions and become able to use them as well as to help them deepen their understanding of multiplication and division in general. *(Japan; 6th grade numbers and calculations objective)*

Notice the natural progression of knowledge required by these Japanese standards with regard to decimals and common fractions. In the third grade, students are required to use decimals and common fractions to represent sizes of quantities. In the fourth grade, students must be able to add and subtract using decimals and common fractions. In the fifth grade, students must be able to multiply and divide using decimals. In the sixth grade, students must be able to multiply and divide using common fractions. These requirements are, of course, spelled out in more detail in the content sections of the Japanese standards, but these statements of objectives amply demonstrate the cumulative nature of the knowledge that Japanese teachers must pass on to their students. Each requirement is based on the knowledge learned in earlier grades.

At each stage of their education, then, students in Japan have the foundation of knowledge they need to receive new information, make sense of it, and use it to carry out new computational activities. On this point, as in so many other areas, California's math standards are very similar to Japan's. If one examines the number sense sections of the California math standards in grades three to six, one finds almost exactly the same logical progression of knowledge in regards to decimals and fractions.

Assessment

Many people think that once a state has adopted a good set of content standards, then that's the end of it. Crafting the content standards, however, is just one part of a comprehensive system. Without an assessment device, there will be no way of knowing if the content standards are being met in the classroom. If the assessment device does not accurately measure the knowledge content, then it will also be impossible to determine if the standards are being met. A good assessment device, therefore, is vital to a good standards system.

NAEP Assessment Development Process

One of the most well-known and well-respected assessments in the country is the National Assessment of Educational Progress (NAEP) which is overseen by the National Assessment Governing Board (NAGB). The NAEP tests in math and reading are administered in approximately 40 states and are one of the key instruments used to compare the achievement of students across the nation. It is, therefore, useful to understand the process that the NAGB goes through in order to develop the NAEP test.

According to Mary Lyn Borque of the NAGB, the NAEP development process includes a number of distinct steps. First, there is the development of assessment frameworks which specify content to be tested in a subject. Next is the development of the test and item specifications which constitute the test blueprint. In her testimony to the California Commission for the Establishment of Academic Content and Performance Standards (hereinafter referred to as the California Standards Commission), Ms. Borque pointed out:

> This is the blueprint which defines exactly what should be in the test; what the kinds of items are; the mixed format of items; whether it will be all performance or a mix of performance and multiple choice; how many there will be; how difficult they will be; how they will be scored; how they will be constructed; how the composite score will be put together; and whether there will be subscales or not.[27]

After the test blueprint is developed, then preliminary performance standards are developed to provide benchmarks that will tell the test contractor and others "how good is good enough in math or in reading or in writing."[28] The testing contractor is then given the content specifications, the testing blueprint, and the preliminary performance standards and is asked to develop test items to measure the content specifications. After the test items are written, the test is administered to students. Once the results are tabulated, then the NAGB convenes panels of experts to make recommendations (after considering all aspects of the exam including multiple choice and open-ended questions) as to the so-called "cut scores," i.e. what scores should distinguish between "advanced" and "proficient" levels of performance or between "proficient" and "basic." Only after all this has been done does the NAGB pull out examples of student work that help exemplify "basic," "proficient," or "advanced" levels of performance to parents and the general public.

What Type of Assessment?

What type of assessment device should a state adopt? Of course the assessment device must be aligned to the content standards. But should the test rely strictly on multiple-choice questions, strictly on performance-based assessments (performance-based assessments are not related to performance standards, but, rather, refer to types of assessments where students are required to construct responses to questions, such as open-ended essay questions, or tasks, such as doing a science lab), or some combination of the two?

Also, one must ask what is the assessment device's purpose? Is it to meet the seeming needs of students and parents (e.g., focusing on individual student work) or is it designed to meet the concerns of the state (e.g., cost and reliability)?

In her testimony before the California Standards Commission, Michigan State University Professor Susan Phillips, one of the nation's top experts on standards and testing, listed a number of critical tradeoffs between multiple-choice questions and performance-based assessments:

- **Depth versus breadth.** "[W]ith multiple-choice items, you can get more individual examples of student knowledge and skills. And the way I like to think about it is you have this big domain of things that you want students to learn. [Y]ou can't test people for seven days in a row, so you're going to make some choices. You can't ask everything that you'd like to know that [students] may have learned. So you choose a sample. And from that sample you want to be able to infer to the bigger domain. You

want that sample to tell you something about all the things that were learned. Now with multiple choice, because it takes less time to respond to an individual item, you can ask more items so your sample of individual things is greater. On the other hand, with the performance type assessment or open-ended item, it takes more time for the student to respond, but you get more depth about the particular thing you asked because the student is asked to provide explanation, examples, usually some kind of written response that tells you more about the thought processes and requires the student to support their answer as opposed to just picking an answer. So you get depth, but you only get it in one particular area. And if you went to all performance assessment, you would have very few specific and individual things that you could measure."[29]

§ **Time and cost of scoring.** "Many states started out thinking that they wanted to do all performance assessment. And part of that was because they believed that they needed that to measure things. I think people have found since then you don't have to have performance to measure certain kinds of skills. You can score multiple-choice [questions] by machine, it's very efficient, it's very cost effective. When it comes to the open-ended items, you have to have trained raters [i.e., scorers] .... That costs lots of money. It's very resource intensive. And so if you had unlimited resources, of course you could do all performance assessment, but most states have limited resources and they have to balance what they can spend for in terms of the scoring versus the amount of information they're going to get by doing that."[30]

§ **Ability to generalize.** "[I]f they are well constructed, you can take two sets of multiple-choice items built to the same specifications, give them to students and get very similar scores.... Performance tasks, on the other hand, don't seem to generalize as well. That is, if you have two tasks that measure the same thought process, the same skill, the same content, students won't always do equally well. In fact, you can get cases where they do well on one and not on the other."[31]

- **Factual knowledge versus higher order thinking skills.** "Another thing I would just underscore is that some people think that multiple-choice items can only measure factual knowledge and basic skills. That's not true. Well written multiple-choice items can measure higher order thinking skills. Likewise, performance assessments, by definition, don't necessarily measure higher order skills. You can, in fact, construct the performance assessment that does very low level things."[32]

- **Memorability.** "Even if they are not released, [performance questions] will be remembered because there are so few [test questions], they're unique, they're distinctive, and so you're going to end up having to replace those even faster than the multiple choice. And again, that's costly to do because the development process is long and involved. You've got not only the items to develop, but you've got the rubrics and then you've got the training of the scorers and developing the work that goes with it."[33]

- **Standardization.** "You have to be fair to everybody. And part of being fair means that you have to give everybody the same opportunity to be successful. That is, they all have to be responding to the same task, the same situation, the same items, whatever it is that you've chosen in a standardized way so nobody has a particular advantage over another."[34] (The standardization problem is particularly apparent in the portfolio performance assessment discussed below).

Other points stressed by experts include:

- **Equivalency.** Assessments from one year to the next must be equivalent in content and equivalent statistically. If they are not, then the results of the assessment can be seriously misleading (i.e., gains or losses on student scores may not truly indicate gains or losses in student knowledge). Performance-based assessments are more susceptible to equivalency problems.[35]

- **Validity.** Do increases or decreases in student scores reflect actual gains or losses in student knowledge and achievement? If yes, then the assessment results

are valid. If not, then the gains or losses are spurious and the results are invalid.[36]

Scoring

Part of the standardization problem involves scoring. It must be pointed out, first of all, that there is some subjectivity in any assessment process. That being said, however, it is still true that some types of tests are more reliable than others, which is especially important in high-stakes accountability systems where serious consequences for schools are linked to the level of student performance on an assessment device. In a high-stakes accountability system, accurate, and unambiguous measurement is critical.

On a multiple-choice question, there is only one right answer so there is no ambiguity about scoring. Not so with performance-based assessments such as open-ended essay questions. According to Brad Thayer of National Computer Systems, the nation's largest test scoring firm:

> If you give an essay [to score] to 100 teachers nationwide and even if you give them a scoring guide, they're going to approach it differently. Some will grade it more heavily on grammar. Others on the content.[37]

Indeed, researchers have found that two trained scorers will agree only sixty to eighty percent of the time when grading an essay using a 1-to-5 scale.[38] Horror stories of idiosyncratic grading by scorers abound:

> Another reader—a single mother putting her daughter through medical school—bristled at instructions for grading a short essay on a 1-to-3 scale. Any response written in capital letters, no matter how brilliant, could not get a 3. But the only way to get the lowest score, a 1, was to write an "unintelligible" answer. The result? Answers differing widely in quality earned a 2.[39]

Perhaps the most difficult type of performance-based assessment with regard to standardization is the portfolio assessment. Portfolio assessments require a student to perform various tasks in class, that is, write essays, do individual projects, participate in group projects, etc. The results of which are then collected and put into a portfolio for that student. In many cases, the teachers and students are unrestricted in the selection of tasks and the work products that result from those tasks. These tasks and work products, thus, vary greatly from school to school. Portfolios are then evaluated and scored. According to Professor Phillips:

> Portfolios are a wonderful tool in the classroom and it's a wonderful way for teachers to show parents at parent conference, for example, what's been going on in the classroom, what students have been doing. It's very, very difficult to implement that process at the state level to use in some kind of high-stakes accountability process. There are too many things you cannot control. You can't standardize the conditions under which the [students'] work was developed. You don't know whose work it really is. You don't know how much help the student got, if they worked in groups. You don't even know that that student made a substantial contribution. The student may have simply taken the work of others.[40]

In 1990, Kentucky passed an education reform act that required portfolio assessment of students, but an academic review panel found huge problems in the system. According to the panel's findings, large scoring gains by Kentucky students on the portfolio assessment failed to be matched by similar gains on other tests such as the National Assessment of Educational Progress, which strongly suggested that the test results were invalid. Also, use of ad-hoc judgmental procedures for linking or equating assessments from year to year likely resulted in making the year-to-year comparison of scores invalid. Further, because teachers in the students' own schools scored the portfolios, results were biased upward by a significant amount. There were also no control factors affecting reliability such as initial instructions from teachers; the amount and type of pre-teaching using similar tasks; time allotted to performance of the tasks; amount and type of assistance from teachers, peers, parents, and others in performing the tasks; opportunities for revision; and the amount and type of assistance in revision provided by others. Because of these problems, Kentucky has reinstituted a standardized norm-referenced test.[41]

A similar Rand study of Vermont's portfolio assessment system came to virtually the same conclusions. According to the Rand researchers, scorers were confused by the scoring rubrics (i.e., guidelines) and disagreed amongst themselves about scoring details. Training a large number of scorers so they could score portfolios accurately also proved to be very difficult. Lack of standardization of tasks, especially in subjects such as math and science where tasks can vary greatly, also impeded reliable scoring. There were also problems with validity, and serious generalizability problems because the sampling of tasks varied greatly from one classroom or school to another. Finally, financial costs of the system were astronomic.[42]

As mentioned earlier, one of the most crucial fallouts of poor assessment systems, such as those emphasizing portfolio assessments, is the domino effect felt throughout an entire school accountability system,

especially a high-stakes accountability system where rewards and sanctions for schools are based on assessment results. If the assessments are unsound, then the classification of students and schools will be unsound, which could result in the triggering of inappropriate rewards or sanctions.

The bottom line is that multiple-choice questions offer the best value in terms of cost, breadth, reliability (i.e., accuracy with low measurement error), validity, and ability to generalize. In its review of the Kentucky performance-only assessment system, the academic panel strongly recommended the re-incorporation of multiple-choice questions, saying that their previous elimination resulted in serious consequences for the state testing system:

> The elimination of multiple-choice items by the [Kentucky] Department of Education from all of the important analyses unnecessarily restricts content coverage, lowers the reliability of the school accountability index, reduces the stability of the equating or linking of assessments from one year to the next, reduces the stability of the performance standards, and creates less reliable and valid scores for school and individual score reporting.[43]

All of this is not to say that performance-based assessments cannot be used at the local or classroom level. They certainly can. The state, however, because it must implement an accountability system, must focus more on issues of reliability. Performance-based assessments, therefore, should be limited to areas where multiple-choice questions are simply inappropriate. As Susan Phillips prescribes:

> So that's why it can be optimum to use multiple choice where that's appropriate to get multiple samples of behavior, and use the performance assessment where the objective standard calls for something that cannot be demonstrated other than by a performance.[44]

Score Reporting

The way scores are reported rivals in importance the kind of assessment in use. Are individual student scores reported or are only score averages reported for schools or school districts? Will matrix sampling be used?

The major plus for matrix sampling is that it allows for very wide coverage of content. Since there is only a limited amount of time available for testing, it is often the case that a single assessment document cannot cover all the material in the content standards. In order to widen

this coverage, matrix sampling assembles different assessment documents covering different aspects of the content standards. These are administered to different sample sets of students (in other words, not every student answers the same questions), and aggregate scores are then calculated. The NAEP, for example, uses a matrix sampling method.

The drawback for matrix sampling and other scoring methods that do not report individual student scores is that students often lack motivation to do their best work. Without the reporting of individual scores, students may not work as fast nor answer difficult questions. Parents have also been known to hold their children out from taking the tests. This slacking on the part of students could have serious consequences in a high-stakes accountability system where rewards and sanctions for schools and school personnel are based on assessment results. It should be noted, though, that matrix sampling in a high-stakes accountability system would still motivate teachers to feel responsible for conveying all the information required by the standards since they would not know upon which portion of the standards their students will be tested.

Reporting individual student scores has several advantages. It motivates students to do their best on the test. It informs parents as to how well their children are performing. Informed parents then become a source of pressure on schools to improve. The drawbacks include some added cost, privacy concerns, and less content coverage, at least versus matrix sampling. It is up to state officials which trade-offs they are willing to accept.

Student Work Samples

Finally, there is the question as to whether to first develop an assessment device or to first collect student work samples that represent the various levels of achievement/knowledge/performance. This work product represents what a student with "proficient" or "basic" knowledge would produce, which can then be used in developing the assessment device.

Since collection of samples of student work is often linked to performance standards, putting the collection of student work ahead of the creation of an assessment device effectively means that one is crafting performance standards before the development of the assessment device. (Performance standards are the benchmarks of achievement, aligned with the assessment device, that indicate the degree to which a student has met a content standard. These benchmarks are often designated by terms such as "advanced," "proficient," and "basic" knowledge. A full discussion of performance standards follows in the next section.)

One school of thought argues that collection of student work product must occur early in order to inform education and testing officials as to what constitutes "advanced" knowledge, "proficient" knowledge, or "basic" knowledge of a certain content standard. Samples of student work representing each benchmark (advanced, proficient, etc.) are desig-

nated and these samples then assist those crafting the assessment device. As a memo by the California Standards Commission observed:

> Samples of student work, for example, illustrating the kind of work children should be able to do at the various levels, may be helpful to illuminate the performance levels. They may provide guidance to students, teachers, and parents about "how good" is really "good enough."[45]

One group of education researchers explicitly defines performance standards as including student work samples:

> Performance standards are, therefore, made up of a combination of performance descriptions, work samples, and commentaries on the work samples: the performance descriptions tell what students should know and the ways they should demonstrate the knowledge and skills they have acquired; the work samples show work that illustrates standard-setting performances in relation to parts of the standards; the commentaries explain why the work is standard-setting with reference to the relevant performance description or descriptions.[46]

There are, however, very serious problems with gathering student samples first and then keying the assessment device to them. In her testimony to the California Standards Commission, the NAGB's Mary Lyn Borque pointed out three major difficulties:

- **Representative work.** "[I]f your standards are being set to raise the high bar ... then there's no guarantee that looking within the current performance distribution you're going to find student work that represents work at or above the high bar . . ."[47]

- **Content domain.** "Second, it could limit the domain of content. Now what do I mean by that? Let me give you an example. Suppose that one of the goals of the California Writing Program is to teach persuasive writing starting as early as grade three or four and moving up through the higher grades, up through the high school. And suppose you go out and you try to find pieces of persuasive writing at grades three or four, yet no teacher in California might be teaching persuasive writing. So if you only gather from what is out there, you're

going to not only limit the performance distribution, but you're also going to limit the content domain."[48]

§ **Reliability and validity.** "And then the third disadvantage, of course, is that it's clear in [California's] legislation that you need to set standards that are reliable and valid. And it does cast doubt on the validity of standards when you focus excessively on where standards are now and not on where students should be."[49]

Susan Phillips cautions that collecting student work samples and basing assessment on those work samples raises so-called elicitation problems (i.e., problems regarding how the work samples are collected). Addressing the California Standards Commission, Professor Phillips, like Ms. Borque, warned about the representativeness of student work with regard to new higher standards:

> Particularly in terms of what I've read about [California's] standards, you have statements in writing that say that you're aiming at a world class standard, that you think you're not there yet. And so I think there's a very real possibility that you wouldn't elicit your top level behavior necessarily in going after the work on the front end.[50]

Further, says Professor Phillips, there are problems involving standardization (i.e., making sure that test conditions are the same for all students):

> [Y]ou're going to have to be very careful about what it is that causes the student to do the performance that you're scoring and you're going to want those conditions to be the same for everybody. And if you just go around and collect student work at some point, one of the things you don't know is what elicited that work. In other words, what background information they had, what directions they were given, how much assistance they were given, how much technology and other kinds of things they could use to put it together.[51]

In addition, the process of gathering student work samples often does not control for the time given to students to accomplish assigned tasks. Thus, a student work sample judged to be representative of "advanced" knowledge may be tainted because the student author of

the sample may have had more time to work on the assignment than students in other schools.

What these comments seem to say is that while the names of performance levels (e.g., "advanced," "proficient," "basic," etc.) may be adopted at any time, either before or after adoption of an assessment device, gathering and using student work samples to give meaning to those levels prior to the development and administration of the assessment device is unwise. Student work samples may be valuable in other documents such as teacher guides or curriculum frameworks, but they should not be the key factors in the standards process.

Performance Standards

Performance standards are often defined generally as benchmarks which describe "how good is good enough," i.e. the knowledge a student must demonstrate in order to show that he/she has a certain level of understanding of a content standard (in other words, performance standards make content standards operational). According to the California Standards Commission, this common definition can be stated as follows:

> One definition is "achievement levels," meaning that levels of achievement are laid out, and meant to indicate the degree to which a student has met a content standard. NAEP's levels are a good example ("basic," "proficient," "advanced," and an implied fourth level in which the student is "below basic"—or not on the chart at all). The [California] Academic Standards Commission recommended five levels: Merit, Proficient, Nearly Proficient, Below Proficient, and Well Below Proficient. Even traditional grades (which theoretically encompass a range of "scores") do the same thing.[52]

The U.S. Department of Education gives a similar definition of performance standards, but supplements the definition with descriptions of what the student must know and do to achieve the various performance levels:

> [Performance standards] flesh out content standards in two ways. First, they provide descriptions (and sometimes examples) of what students are expected to know and be able to do to demonstrate that they have reached specific proficiency levels in the knowledge and skills framed by the content standards. Second, performance standards identify explicit levels of achievement in each subject matter set out in the content standards. Performance standards set the categories of proficiency for students and allow a judgment of progress to be made for individual students, for schools, and for larger systems.[53]

It is important to note, however, that there is still no consensus in the education community as to the precise definition of a performance standard. A recent report prepared for the U.S. Department of Education and the Council of Chief State School Officers noted that different groups of people defined performance standards differently:

> To test developers and psychometricians, performance standard usually refers to the point on a test score scale that separates one level of achievement (e.g., pass) from another (e.g., fail), identified through a technically sound process. To educators involved in the development of curriculum and instruction, performance standard often means a description of what a student knows and can do to demonstrate proficiency on a content standard or cluster of content standards. To others, the term performance standard indicates examples of student work that illustrate world-class performance.[54]

The practical reality, however, is that once the scores from the assessment are available, most people are interested in knowing how students performed. The achievement levels and their alignment to the test scores tell people how well students fared. Thus, the widely used definition involving achievement levels should be the operational definition that should guide the crafting of any performance standards.

It is important to note the quality interdependence of the content standards, the assessment device, and the performance standards. According to Linda Bond, director of assessment at the federal North Central Regional Educational Laboratory (NCREL), most states first set their content standards, next develop their assessments, and then establish scoring levels (cut scores) that serve as performance standards. The result is that "the quality of performance standards (i.e., cut scores) is dependent upon how well the important content standards, and the level of performance deemed satisfactory on those standards, are reflected in the content of the test."[55]

Based on the NAGB's procedure for developing performance standards for the NAEP and recommendations of the California Standards Commission staff, here are the important steps in crafting a set of performance standards:

§ **Set the number of performance levels.** As indicated above, the NAEP has three levels (with an implied fourth), while the California Standards Commission has proposed five. The NAEP performance standards were designed so as to produce a bell curve with a top, a middle, and a bottom. It is probably best to have between three and five levels. Having a

single level would limit the amount of information about the degrees of student knowledge that could be gleaned from the test results. For example, if the one level was "pass," it would be impossible to say how many students had outstanding levels of knowledge versus those who had barely an adequate amount. On the other hand, too many levels would make it difficult to discriminate between one level and another because, according to Mary Lyn Borque, "the more of these cut points that you have on a distribution, the closer they get [to each other]."[56]

§ **Name the levels.** The NAEP uses the terms "advanced," "proficient," and "basic," with an implied category of "below basic." Some states have shied away from designations that imply that students have failed. The problem with this view is it risks sugar-coating unpleasant facts. In Kentucky's performance levels, the lowest level is termed "novice" which means, among other things, that a student shows minimal understanding of core concepts. Does the term "novice" convey this low level of understanding? A harsher term may be more accurate, may inform the public better, and may create greater pressure for schools to improve.

§ **Provide descriptions of content and quality of performance at each level.** NCREL's Linda Bond says that matching test content with content standards "can be further improved by providing assessment developers with a written description of what satisfactory student performance looks like—performance descriptors."[57] The performance descriptors are developed before assessments, ideally at each grade level. Observes Ms. Bond, "Consideration should be given to including sample tasks that illustrate the kinds of things a competent student knows, understands, and can demonstrate."[58] (Although some argue that these descriptors cannot be made meaningful without student work samples to guide the process, the reality is that many states and localities have devised descriptors without resorting to work samples.) The following is NAEP's fourth-grade "proficient" performance description for writing:

> Fourth-grade students performing at the proficient level should be able to draft an effectively organized response that shows a clear understanding of the writing task they have been assigned. Their writing should include details that support and develop the main idea of the piece, and its form, content, and language should show that these students are aware of the audience they are expected to address. The grammar, spelling, and capitalization in the work should be accurate enough to communicate to a reader; there may be some minor mistakes, but these should not get in the way of meaning.[59]

It should be pointed out that some states refer to their performance descriptors as their performance standards. However, performance descriptors are only a part of a performance standard. The performance standards should also include the cut scores on assessments which are necessary to decide how well a student has performed. (discussed below)

- **Test items are developed and administered.** As in the earlier discussion of the NAEP process, the test contractor takes the content specifications, the testing blueprint, and the preliminary performance standards and comes up with test items which are administered to students (the test items could be sample items).

- **Cut scores decided.** According to Linda Bond, "When assessments have been developed and administered, the next issue becomes one of determining passing levels."[60] These passing levels are the cut scores for the test that demarcate the boundary between the various achievement levels. There are a number of statistical methods for determining and setting these cut scores including predetermining scale points as performance levels (the NAEP method), predetermining performance levels and finding scale points to match, the Angoff method, the Nedelsky method, the contrasted groups method, and the norm-referenced cut method.[61] Some of these methods are better than others (for example, the Angoff and Nedelsky methods both have reliability problems and the norm-referenced method has been criticized as an arbitrary approach).

§ **Student work samples provided.** Samples of student work exemplifying performance on the assessment device at each level are chosen and released.

Once the number and names of performance levels have been decided, performance descriptors have been issued, the test developed and administered, and cut scores have been picked, then one has a performance standards system that puts the content standards into operation.

Implementation and Accountability

Even after a good set of academic standards have been adopted, after a good assessment device has been developed, and after good performance standards have been crafted, all could be for naught unless a state has an effective implementation and accountability mechanism (currently, many states that have content and performance standards do not have accountability systems). Such a mechanism will be needed because many school and district officials will often balk at implementing rigorous content standards since failure of substantial numbers of students to meet those standards will reflect badly on the job performance of teachers and administrators.

A good assessment device, aligned to the content standards, will provide significant incentive for schools to implement the content standards. This is so because unless the content standards are implemented, students will most likely do very badly on an assessment device aligned to the content standards. Such a result would not be good news for teachers and local school officials. One can see, therefore, why aligning the assessment device with the content standards makes both educational and practical sense. There should be other incentives, however.

Some states have set up various kinds of accountability systems that are meant to encourage schools to implement and meet the content standards. These accountability systems fall roughly into two main types: high-stakes accountability and low-stakes accountability. High-stakes accountability systems involve the use of rewards and sanctions on students and schools for meeting or not meeting the content standards. Low-stakes systems are, basically, all other systems that do not involve rewards and sanctions.

If states are really serious about implementing their content standards, then a high-stakes accountability system is strongly recommended. Under a high-stakes system, the rewards and sanctions focus on either students, schools, or both. The following are examples of high-stakes accountability systems in a number of states:

- **Virginia.** Starting in 2004, 12th-grade students must pass a series of content-based exams in order to graduate. Also, 70% of students in a school must pass the exams, or schools could lose their accreditation.

- **Arizona.** Students will need to score at a "proficient" level in order to graduate from high school.

- **Kentucky.** Kentucky uses its performance-based assessment system to establish scores for individual schools which are, in turn, used to set improvement goals. Schools exceeding their goals are eligible for financial rewards, while schools not meeting their goals are assigned state managers. There is no high-stakes accountability for students. As mentioned previously in this paper, the combination of Kentucky's unreliable assessment system with its high-stakes accountability system has yielded disastrous results. The academic panel reviewing Kentucky's assessment and accountability system concluded: "The misclassification of schools in some reward categories are high and, therefore, the rewards and sanctions may be difficult to defend."[62] The lesson in Kentucky is that a high-stakes accountability system must be accompanied by good standards, a good assessment device, and good performance standards.

With regard to high-stakes student accountability, the American Federation of Teachers (AFT), which has been in the forefront of the movement to promote tough and rigorous content standards, says:

> Learning complicated material requires diligent studying and constant practice, which students won't undertake unless there are clear, significant incentives for doing so. Incentives should include access to higher education, training, and jobs, but they should also include more immediate rewards, such as prestigious citations, special trips, and scholarships—and more immediate consequences, such as required summer and weekend catch-up classes (which would also signal that they might as well learn the material the first time, since eventually they will have to learn it).[63]

The AFT also advocates ending automatic promotion of students to higher grades if students fail to meet academic content standards. Students in danger of being retained or who have been retained should,

according to the AFT, be provided with intensive tutoring or other special instructional assistance.[64] California has just enacted a law that ends social promotion.

In addition, many education experts recommend the use of exit exams for high school seniors. Exit exams test students on their cumulative competency in core subjects such as math, science, and English language arts. Students must receive passing grades on the exams in order to graduate. The exams could be developed as part of the assessment crafting process associated with a standards system and could be aligned to the standards.

Exit exams have much to recommend them. First, such exams would give real incentive for students and schools to meet the content standards. Second, by requiring students to demonstrate their competency in core subjects, exit exams would give real weight to the high-school diploma. No longer would employers have to wonder whether a high school graduate could do elementary math or read a manual. Third, exit exams would inform the public as to the progress of student performance. Obviously, if exit exams are to be implemented, then individual scores on exams would have to be reported. Although the drawbacks of individual score reporting would apply, the benefits would still likely outweigh such concerns.

Although linking student performance to the content standards is an important incentive for students to take the standards seriously, it is equally important to provide incentives for schools and school personnel as well. It does no good to have motivated students if teachers and principals are unmotivated. Schools and school personnel, therefore, must be subject to rewards and sanctions if a high-stakes accountability systems is to succeed.

Many of the rewards and sanctions proposed for schools and school personnel, however, do not create real incentives for schools and school personnel to improve their performance. Take, for example, the proposed rewards and sanctions contained in the report of the California Department of Education's Rewards and Interventions Advisory Committee (RIAC). As in other states, RIAC recommends the usual set of monetary and nonmonetary rewards for schools that meet short-term and long-term performance targets, based upon statewide assessments and the state's content standards. Schools that do not meet these short-term and long-term performance goals would be required to develop a school action plan that focuses on student achievement, with those scoring below average on the state performance standards being assigned to a three-phase intervention's process.[65]

The three-phase intervention's process is instructive. In the first phase of the RIAC intervention process, low-performing schools would hire "academic coaches" to assist principals in developing and implementing the school's action plan, and would receive extra funding from the state and the school district to implement the strategies contained in

the action plan. In the second phase, schools falling short of their short-term performance targets after two years would receive state and school district funding to hire a school improvement expert and implement strategies for improving student academic achievement. Finally, in the third phase, schools that still fail to meet their short-term performance targets after another two years would either be required to continue phase two activities, or be subject to a state takeover. This would include reassignment or transfer of students, teachers, or other school staff and the reallocation of school resources, closure, or other action.[66] RIAC says, however, that their goal is "not to close schools or terminate teachers."[67] There are several critical problems with RIAC's three-phase intervention's process.

First, in phase one and phase two, schools that perform badly are not sanctioned, but are actually rewarded with more state and local district funding. This creates a perverse incentive for schools to not achieve performance targets. Further, what if the reason for a school performing badly is the fact that the school has low-quality teachers and administrators? During phases one and two, nothing could be done about that. That means that for at least four years, students may be subjected to continued low-quality instruction. Indeed, a high school class could graduate during this phase one and two period. Even during phase three, poorly performing teachers and administrators are not fired, which is in accord with RIAC's goal of not terminating teachers. They are merely reassigned to impose their poor quality work on students in other schools. Unless such poorly performing personnel are faced with termination, which means the elimination of the tenure system, there is little real incentive for them to improve their performance. Also, reassignment of poorly performing personnel makes closure of a poorly performing school a hollow gesture, and RIAC says their goal is not really to close down schools anyway.

As one can see, if states are unwilling to combine real sticks with their monetary carrots then high-stakes accountability systems will be destined to come up short in delivering the performance results promised by the content standards. There are, however, more promising approaches to creating real incentives for schools and school personnel.

In *Making Schools Work*, edited by top education economist Eric Hanushek, the Brookings Institution outlines an innovative incentive system that, upon reflection, could be tied to and become the basis for a high-stakes accountability system:

§ **Performance contracting.** The basic idea of performance contracting, says Brookings, is that instead of school districts employing teachers and administrators directly, districts instead contract with independent firms to provide educational services, such as teaching, at schools.[68] Since the terms of the con-

tract could include any provisions mutually agreeable between the district and the firm, one contract requirement could be that the firm implement and adhere to the content standards. Also, the firm could be required to achieve specified improvements in student achievement based upon results from the standards-aligned assessment. In other words, the contractor would be paid according to outcomes. Such performance contracting could either be instituted at all schools or only at badly performing schools, although optimally it should be an option available to all schools. Performance contracting is an attractive solution if a school's problem is low-quality school personnel.

§ **Merit pay for teachers.** Merit pay links teacher pay directly with performance. Rewards, says Brookings, are based on results rather than behavior "so they circumvent the difficulties in defining a *priori* what good teachers or good teaching might be."[69] Under a merit pay system, teacher pay could be linked with the assessment results of students. A teacher's base pay could be supplemented with bonuses based upon how well the teacher's students perform on the state's assessment device. In California, the state Legislative Analyst's Office has recommended implementation of merit pay demonstration programs tied to, among other things, student achievement on the standards-aligned assessment.[70] Merit bonuses could take the place of the automatic "step raises" that teachers get simply for remaining on the job.

§ **Teacher selection and renewal.** Brookings points out that private schools directly link teacher hiring/selection procedures and teacher retention policies with classroom performance.[71] Teachers should be under a performance contract whereby their continued retention is based to a significant extent on their implementation of the content standards and the performance of students on the standards-aligned assessment device. Although a teacher could be given opportunities to improve through training courses, etc., those not performing at acceptable levels would be terminated—a fate, observes Brookings, "that, today, befalls only the most grossly and demonstrably incompetent."[72] Once again, this

reform would require an overhaul of current tenure systems which means reforming the collective bargaining process between school districts and teachers' unions. Given the reality that such an overhaul would be very difficult, reformers should start with the more feasible task of requiring performance contracting for newly hired public school teachers.

§ **School choice.** Brookings notes that, "Most public schools effectively have a monopoly; parents living in a certain area have no choice over which school their children attend."[73] Giving parents and students voucher/opportunity scholarships that can be used at the private or public school of their choice breaks this monopoly. More important, says Brookings:

> Allowing students to choose which school to attend is meant to encourage them to attend better schools. That is a particularly valuable opportunity in inner cities, where families frequently lack the resources to move to affluent suburbs where good schools are more prevalent. In turn, consumer choice would pressure the poorer performing (unpopular) schools to improve. Giving students and their parents a choice would thus place greater incentives on performance, because students—and presumably resources—would migrate from poor schools to good ones and force all of them to respond to the concerns of parents and to issues of quality.[74]

Although not mentioned in the Brookings Institution book, there are two additional components to any successful implementation system:

§ **Teacher training.** As was mentioned in the introduction to this paper, standards do not, by themselves, solve the problem of teacher incompetency. It is a sad fact that many university schools of education and other teacher training programs issue teacher credentials without requiring that prospective teachers achieve competency in subject-matter content areas. Instead, prospective teachers take course after course on teaching methodology. Newly minted teachers, thus, often know a great deal about how to teach, but little about the content they are supposed to teach. To rectify this situation, two reforms should be instituted.

First, the coursework at state-supported schools of education should be changed to include a solid number of content-based courses which are aligned to the K-12 standards (i.e., that emphasize the knowledge required by the standards that teachers will have to convey to students). Second, entry exams for teachers should be created that measure content knowledge and which are aligned with the K-12 standards. Texas, for example, has put in place a high-stakes accountability system for teacher competency. Under the Texas system, colleges that have below a 70% passage rate on the state's teacher licensure exam lose the right to prepare teachers. Such sanctions should encourage higher education officials to start producing teachers who will champion tough standards rather than run away from them.

§ **Professional development for existing teachers.** Most states have continuing development programs for teachers. Once rigorous content standards have been adopted, these programs should be reformulated to coach teachers on the content they need to impart to their students. Teachers must be just as prepared as their students to handle the new standards. Kati Haycock, executive director of the Education Trust, notes that research from New York, California, and Texas shows that professional development programs that focus on the content that students must learn have succeeded in raising student achievement.[75]

Regarding the school choice recommendation, it is important to remember that the Brookings book was published in 1994, when only fragmentary evidence existed on the effectiveness of school choice. Since then, more and powerful data has been collected showing that school choice does work to improve student performance.

Studies of the Milwaukee school choice program by researchers from Harvard University and the University of Houston found that students attending private school through the school choice program performed better on math and reading tests when compared to a control group of students in the public schools.[76] In addition, the Wisconsin Supreme Court has just upheld the constitutionality of the Milwaukee program providing a major legal victory for school choice proponents. The evidence from overseas is also compelling.

According to a study by the London School of Economics, students attending private school through the British school choice program scored much higher on college entrance tests than a control group of their peers .

in government-run schools.[77] Finally, rather than hurting the quality of public schools, Harvard economics professor Caroline Minter Hoxby has found that increased competition actually benefits the public schools.

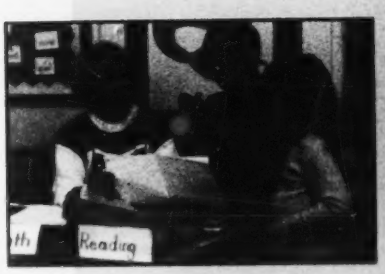
According to Professor Minter Hoxby, "A $1,000 voucher would improve student performance across the board: both public and private school students would increase their educational attainment (about two years), test scores (about 10 percent), and wages (about fourteen percent)."[78] As Professor Minter Hoxby notes, "public schools that face a disproportionate increase in competition because of the vouchers will disproportionately improve their productivity."[79] School choice options could be included in high-stakes accountability programs in a number of ways.

Students at schools designated as low-performing (based upon standards-aligned assessment results and performance standards) could become eligible for vouchers/opportunity scholarships. California Governor Pete Wilson has been pushing such a proposal for the past several years. Although it is imperative that the government keep its regulation of private schools to a minimum, the state could attach a condition to the voucher/opportunity scholarship that the scholarships could only be used at private schools that adhere to the academic content standards and participate in the standards-aligned assessment. This would have the advantage of giving parents more information upon which to make their decision as to what school to send their child (it also addresses the oft-repeated concern that parents do not have enough information to make wise choices).

Private schools that do not want to accept that condition are free to opt out of participation in the school choice program. Also, private schools that do accept the scholarships should have their freedom to innovate (e.g., freedom to hire/fire teachers, freedom to decide upon teaching methods, freedom to contract out services, etc.) protected from government intrusion. Legislative supermajorities could be required for any added regulation of private schools. In the end, school choice offers students, especially those trapped in poorly performing public schools, the opportunity to receive a better education. It also gives public-school teachers, administrators, and district officials a big incentive to improve their performance or else risk losing students and the per-pupil funding that comes along with them.

If a high-stakes accountability program is to incorporate these much-needed reforms, then there will have to be dramatic changes in state and local district rules, regulations, and procedures. The issue of collective bargaining, for example, would have to be rethought. It is always easier to tinker around the edges of the current system by increasing funding, moving around staff, etc., than to change the system in a fundamental way. Yet, if the system itself is the reason for the current low state of performance, then fundamental changes are necessary. Indeed, it is only through such fundamental changes, tied to an accountability program, that the goal of standards—improved student knowledge and achievement—will become a reality.

Communications



All of the effort to craft content standards, assessment devices, performance standards, and implementation and accountability programs will be for naught if the details of these systems are not disseminated to parents, teachers, school administrators, and the general public. It is important to remember that this is not simply an exercise for education insiders in state capitals. If people at the local level are made aware of the standards and come to support their implementation, there is a good chance that the standards will succeed. It is critical, therefore, that state education and elected officials come up with a comprehensive communications plan to inform people about the standards system.

An effective communications plan should include several important elements. First, classroom teachers should be made fully aware of the details of the standards and the various programs attached to them. Too often teachers are at the bottom of the trickle down of information. The fact is, though, teachers are usually the first and best source of information for students and parents. Keep teachers informed and it is very likely that students and parents will become informed as well. Teacher guides, professional development seminars, regular meetings with state and district officials, and newsletters are just a few of the ways that teachers can be kept in the loop.

Another important communications tool, of course, is the media. Doyle and Pimental recommend regular ongoing meetings with reporters and editorial board members to keep them informed. Given that standards, assessments, etc. are complex issues, a great deal background information must be made available to reporters so they can provide full and comprehensive coverage.

Most important, however, is parental involvement. Written explanations of the standards and assessments should be given to all parents in easy-to-understand booklets, flyers, brochures, etc. Parents must be aware of what is expected of their children and the stakes involved. Only then can parents become allies in ensuring that their children meet the new standards and become watchdogs over the schools to see that teachers and administrators are doing their jobs adequately. When both parents and teachers have a full understanding of the requirements of the standards, then they can work together through parent-teacher conferences, parent-teacher organizations, and other joint activities to make the

standards a reality in the classroom.

Finally, the rest of the community (business, organizations, churches, etc.) should not be overlooked. For example, given the importance of an educated workforce to employers, businesses and business groups should be fully informed about the importance of a comprehensive standards system. Businesses could then become partners with schools in making sure that students meet the standards.

Summary

This standards template has covered a lot of ground in the preceding pages. The issues involved have been complex and often subtle. A recap, therefore, is in order.

First, a good set of academic content standards should be:

1) rigorous,
2) intelligible,
3) measurable,
4) specific,
5) comprehensive,
6) academic,
7) balanced,
8) manageable, and
9) cumulative.

In deciding upon an assessment device, policymakers should bear in mind issues such as:

1) depth vs. breadth,
2) time and cost of scoring,
3) ability to generalize,
4) factual knowledge vs. higher order thinking skills,
5) memorability,
6) standardization,
7) equivalency, and
8) validity.

Given these considerations, multiple-choice questions should be emphasized with performance-based questions reserved for those areas where multiple-choice questions are clearly inappropriate.

In crafting performance standards, the following steps should be observed:

1) set the number of performance levels,
2) name the levels,
3) provide content and quality of performance at each level,
4) test items are developed and administered,

Summary

5) cut scores decided, and
6) student work samples are provided.

In order to ensure that standards become a working reality in the classroom, policymakers should consider these implementation and accountability mechanisms:

1) performance contracting with outside firms to provide educational services,
2) merit pay for teachers,
3) teacher selection and renewal based on performance,
4) school choice,
5) improving teacher training programs by increasing content-area requirements, and
6) improving professional development for existing teachers.

Finally, a communications plan must be formulated that informs parents, teachers, local school officials, the media, and the general public about the details of the standards, assessments, performance standards, and accountability mechanisms.

If standards are to achieve their goal of improving student performance, all these components must be present. As Denis Doyle notes:

> A new triptych must emerge: Standards set, standards met, consequences. All healthy organizations have standards for performance; the standards are subject to even-handed measurement, and the organization is held to them. It is at once that simple and that demanding.[80]

Doyle also notes, however, that public schools are supremely lethargic organizations that are difficult to change from either the inside or the outside. This lethargy stems principally from the fact that public schools are monopolies protected from competition.[81] Without fear of competition, the incentive to reform is often blunted or non-existent. That is why it is so important to open up the education marketplace to competition through programs such as school choice. The bottom line, then, is that only by making public schools compete will there be any consistent incentive to make standards work.

EndNotes

1 Dennis P. Doyle and Susan Pimental, "Raising the Standard," Coalition for Goals 2000 (Thousand Oaks, CA: Corwin Press, Inc./Sage Publications Co. 1997): 33-40. Doyle and Pimental also say that a standard must be "valuable," however, that category is not included here because valuableness is assumed in many of the other categories.
2 Raimi, Ralph A. and Braden, Lawrence S., "State Mathematics Standards," Fordham Report, Vol. 2, No. 3, March 1998: 55.
3 Doyle and Pimental, "Raising the Standard," 34.
4 Raimi and Braden, "State Mathematics Standards," 25.
5 Munroe, Susan and Smith, Terry, "An Appraisal of Geography Standards in 38 States and the District of Columbia," Fordham Report, Vol. 2, No. 2, February 1998: 63.
6 Doyle and Pimental, "Raising the Standard," 35.
7 Finn, Chester E., Petrilli, Michael J., and Vanourek, Gregg, "The State of Standards," Fordham Report, Vol. 2, No. 5, July 1998: 6.
8 Lerner, Lawrence S., "State Science Standards," Fordham Report, Vol. 2, No. 4, March 1998: 4.
9 Doyle and Pimental, "Raising the Standard," 38.
10 Hirsch, E.D., *The Schools We Need And Why We Don't Have Them* (New York: Doubleday 1996): 14.
11 According to Dr. G. Reid Lyon of the National Institute of Child, Health and Human Development, "Longitudinal data indicate that systematic structured phonics instruction results in more favorable outcomes in reading than does a context-emphasis (whole language) approach." Lyon, G. Reid, Research in Learning Disabilities at the NICHD (Bethesda, MD: National Institute of Child, Health and Human Development 1994): 12.
12 Stotsky, Sandra, "State English Standards: An Appraisal of English Language-Arts/Reading Standards in 28 States," Fordham Report, Vol. 1, No. 1, July 1997: 3.
13 Raimi and Braden, "State Mathematics Standards," 55.
14 Doyle and Pimental, "Raising the Standard," 39.
15 Finn, Petrilli and Vanourek, "The State of Standards," 8.
16 Stotsky, "State English Standards," 3.
17 Munroe and Smith, "State Geography Standards," 64.
18 Lerner, "State Science Standards," 4.
19 Sweetman, J., *The Complete Guide to the National Curriculum: Confidential Two.* (Newton Regis: Bracken Press 1991): 8.
20 Tooley, James, *A Market-Led Alternative for the Curriculum: Breaking the Code,* (London: Tufnell Press 1993): 2.
21 Ribbins, Peter, "Telling tales of secondary heads: on educational reform and the National Curriculum" in Chitty, Clyde, *The National Curriculum: Is it Working?,* (Harlow: Longman Information and Reference 1993): 46.
22 Ibid., 60.
23 "Advisors clash on primary tables," *Times Education Supplement,* 19 September 1997: 6.
24 Raimi and Braden, "State Mathematics Standards," 55.
25 Lerner, "State Science Standards," 4.
26 Stotsky, "State English Standards," 3.
27 Borque, Mary Lyn, transcript of testimony to the California Commission for the Establishment of Academic Content and Performance Standards (hereinafter referred to as the California Standards Commission), 11 March 1998: 39.
28 Ibid.

29 Phillips, Susan, transcript of testimony to the California Standards Commission, 11 March 1998: 61-62.
30 Ibid., 62.
31 Ibid., 63.
32 Ibid.
33 Ibid., 64.
34 Ibid., 66.
35 Hambleton, Ronald K., Jaeger, Richard M., Koretz, Daniel, Linn, Robert L., Millman, Jason, and Phillips, Susan E., "Review of the Measurement Quality of the Kentucky Instructional Results Information System, 1991-1994," (hereinafter referred to as "Review of KIRIS"), Office of Educational Accountability, Kentucky General Assembly, 20 June 1995: 4.
36 Ibid.
37 "Can Essay Tests Really Make the Grade?," *Los Angeles Times*, 31 December 1997: A24.
38 Ibid.
39 Ibid, A25.
40 Phillips, testimony to the California Standards Commission, 64.
41 Hambleton, et al, "Review of KIRIS," 5 to 7, and 4-9.
42 Koretz, Daniel, Stecher, Brian, Klein, Stephen and McCaffrey, Daniel, "The Vermont Portfolio Assessment Program," *Educational Measurement: Issues and Practice*, Fall 1994: 12-13. The Rand researchers emphasize a key criticism of performance-based assessments, even those using standardized tasks: "The limited generalizability of performance across tasks we observed in mathematics is the norm in performance assessments, even those that rely on standardized tasks."
43 Hambleton, et al, "Review of KIRIS," 7.
44 Phillips, testimony to the California Standards Commission, 62.
45 "Performance Standards: How Will We Know What Students Know?," California Standards Commission memo, 11 March 1998: 2.
46 National Center on Education and the Economy and the University of Pittsburgh (New Standards), *Performance Standards*, Volume 1, 1997: 9.
47 Borque, testimony to California Standards Commission, 42.
48 Ibid., 42-43.
49 Ibid., 43.
50 Phillips, testimony to the California Standards Commission, 67.
51 Ibid., 76-77.
52 California Standards Commission memo, 29 December 1997, 2.
53 U.S. Department of Education, Draft guidance on standards, assessment, and accountability, (Washington, D.C.: Office of Compensatory Education Programs 1996): 11.
54 Hansche, Linda N., "Handbook for the Development of Performance Standards: Meeting the Requirements of Title I," prepared for the U.S. Department of Education and the Council of Chief State School Officers, Summer 1998: 7.
55 Bond, Linda, "From Content Standards to Performance Standards and Assessments," unpublished paper, 3. Copy available through Pacific Research Institute.
56 Borque, testimony to the California Standards Commission, 35.
57 Bond, "From Content Standards to Performance Standards and Assessments," 3.
58 Ibid., 12.
59 National Assessment Governing Board, "Recommended Final Achievement Levels Description for Writing, Grade 4."
60 Bond, "From Content Standards to Performance Standards and Assessment," 8.
61 Ibid., 8-10. Ms. Bond's paper contains a full discussion of the various statistical methodologies.
62 Hambleton, et al, "Review of KIRIS," 5.
63 "A System of High Standards: What We Mean and Why We Need It," *American Educator*, Spring 1996: 23.
64 Ibid., 26.
65 California Rewards and Interventions Advisory Committee, "Steering by Results: A High-Stakes Rewards and Interventions Program for California Schools and Students," California Department of Education, 1997.
66 Ibid., 11.
67 Ibid., 31.

68 Hanushek, Eric, ed., *Making Schools Work*, (Washington, D.C.: The Brookings Institution 1994): 91.
69 Ibid., 95.
70 Legislative Analyst's Office, Overview of 1998-99 May Revision, Sacramento, CA, 18 May 1998: 12.
71 Hanushek, *Making Schools Work*, 97-98.
72 Ibid., 98.
73 Ibid., 103.
74 Ibid., 104.
75 Haycock, Kati, "Good Teaching Matters," *Thinking K-16*, Vol. 3, Issue 2, Summer 1998:12.
76 Greene, Jay P., Peterson, Paul E., and Du, Jiangtao, "Effectiveness of School Choice: The Milwaukee Experiment," Program in Education Policy and Governance, Center for American Political Studies, Department of Government, Harvard University, March 1997.
77 West, Anne and West, Robert, "Examination Results of Pupils Offered Assisted Places: comparing GCE Advanced level results in independent schools," *Educational Studies* (UK), Vol. 23, No. 2, 1997: 287-293.
78 Minter Hoxby, Caroline, "The Effects of Private School Vouchers on Schools and Students" in *Holding Schools Accountable*, Helen F. Ladd, ed. (Washington, D.C.: The Brookings Institution, 1996): 201.
79 Ibid.
80 Finn, Petrilli and Vanourek, "The State of Standards," 16.
81 Ibid.

About the Author

Lance T. Izumi

Lance T. Izumi is Senior Fellow in California Studies and Co-director of the Center for School Reform at the Pacific Research Institute for Public Policy (PRI), a San Francisco-based public policy think tank. In 1998 he was a Visiting Fellow in Education Studies at the Institute for Economic Affairs, headquartered in London, England. He currently serves as a member of the California Postsecondary Education Commission.

Mr. Izumi has also served as a consultant on welfare reform to the California Department of Social Services, a consultant on juvenile crime to the Governor's Office of Criminal Justice Planning, and as co-chair of Governor Wilson's competitiveness task force on juvenile justice education reform. Mr. Izumi is a columnist and contributing editor for the *California Journal*, the state's premier political monthly. His articles have been published in the *Notre Dame Journal of Law, Ethics and Public Policy, Harvard Asian American Policy Review, National Review, Wall Street Journal Europe, Sunday Times* (of London), *Los Angeles Times, Investor's Business Daily, San Francisco Chronicle, Orange County Register, Sacramento Bee, Los Angeles Daily News, San Diego Union-Tribune*, and many other publications. Mr. Izumi is a regular contributor to the "Perspectives" opinion series on KQED-FM, the National Public Radio affiliate in San Francisco.

Prior to going into the think-tank world, Mr. Izumi served as chief speechwriter and director of writing and research for California Governor George Deukmejian. He also served in the Reagan administration as speechwriter to United States Attorney General Edwin Meese III.

Mr. Izumi received his masters degree in political science from the University of California at Davis and his juris doctorate from the University of Southern California Law Center. He received his bachelor of arts degree in economics and history from the University of California at Los Angeles.

Templates for Legislative and Policy Reform

Developing and Implementing Academic Standards: A Template for Legislative and Policy Reform, by Lance T. Izumi, $16.95

In his standards template, Center for School Reform Co-director Lance T. Izumi provides a practical guide for state lawmakers and education officials who want to craft challenging standards. The template includes an eight-point framework for creating world-class standards and numerous examples of excellent and poor standards worldwide. Also included are sections on implementation, assessment, and performance standards.

Making School Choice Work for All Families: A Template for Legislative and Policy Reform, by John E. Coons and Stephen D. Sugarman, $19.95

UC-Berkeley, Boalt Hall School of Law, professors John Coons and Stephen Sugarman outline legal and political concerns the writer of any universal choice model must consider, describe the many different forms of choice, outline the criteria participating schools should be required to meet, and explain the rights of students who face expulsion.

Expanding the Charter Idea: A Template for Legislative and Policy Reform, by Kenneth Lloyd Billingsley and Pamela Riley, $19.95

Model charter school legislation, a survey and ranking of legislation for each state, and an historical outline of the development of the charter school movement are included in this template by PRI Editorial Director Kenneth Lloyd Billingsley and Center for School Reform Co-Director Pamela Riley. The template assesses charter schools as part of the wider movement toward parental choice in education and offers strategies for further expansion.

To order any of these publications, contact:

PACIFIC RESEARCH INSTITUTE
755 Sansome Street, #450
San Francisco, CA 94111
Phone: (415) 989-0833
Fax: (415) 989-2411
E-mail: pripp@pacificresearch.org
http://www.pacificresearch.org

Nothing contained in this briefing is to be construed as necessarily reflecting the views of the Pacific Research Institute for Public Policy or as an attempt to thwart or aid the passage of any legislation.

HOW TO CONTRIBUTE TO PACIFIC RESEARCH INSTITUTE

❑ $10,000 & above
Sir Antony Fisher Circle

❑ $5,000-$9,999
Chairman's Circle

❑ $2,500-$4,999
President's Circle

❑ $500-$2,499
Patron's Circle

❑ $250 & above
Golden Sponsor

❑ $50
Student

Name
Title/Organization
Address
City/State/Zip
Phone Fax
E-mail

Please make check payable to the PACIFIC RESEARCH INSTITUTE,
a 501(c)(3) organization. All contributions are tax-deductible

❑ I would like to make a contribution of stocks/securities, please call me for details.
I would like to make a contribution of $_____. I would prefer to make
my donation with: ❑ Visa ❑ MasterCard ❑ American Express
Cardholder's Name
Card Number Exp.
Signature

Sir Antony Fisher Circle $10,000 & above
Sponsorship includes: briefing papers, Pacific Clips, President's Message, invitations to Pacific Breakfast clubs, Pacific Forums/lunches, special events, transcripts of events, selected books and fax broadcast information.

Chairman's Circle $5,000-$9,999
Sponsorship includes: Executive Summaries, Pacific Clips, President's Message, invitations to Pacific Breakfast clubs, Pacific Forums/lunches, special events, transcripts of events and selected books.

President's Circle $2,500-$4,999
Sponsorship includes Executive Summaries, Pacific Clips, President's Message, invitations to Pacific Breakfast clubs, Pacific Forums/lunches, special events and transcripts of events.

Patron's Circle $500-$2,499
Sponsorship includes: Executive Summaries, President's Message, invitations to Pacific Breakfast clubs, Pacific Forums/lunches and special events.

Golden Sponsor $250 & above
Special category for contributors of sixty years and over. Sponsorship includes: two free breakfasts for supporter or a friend, Executive Summaries, President's Message, invitations to Pacific Breakfast clubs, Pacific Forums/lunches, special events and discounts on books and briefings.

Student $50
Sponsorship includes: President's Message, invitations to events and student discounts on books and briefings.

Pacific Research Institute
755 Sansome Street, Suite 450
San Francisco, CA 94111
Phone: 415.989.0833
Fax: 415.989.2411
E-mail: pripp@pacificresearch.org
www.pacificresearch.org

IDEAS IN ACTION
PACIFIC RESEARCH INSTITUTE FACT SHEET
January 1999

DEVELOPING AND IMPLEMENTING ACADEMIC STANDARDS: A TEMPLATE FOR LEGISLATIVE AND POLICY REFORM

By Lance T. Izumi

Developing and Implementing Academic Standards is the second in the three-part policy series, *Templates for Legislative and Policy Reform*. The other two templates address the issues of school choice and charter schools. The series recommends best practices and legislative language and highlights problem areas that arise when policy deviates from those best practices.

Developing and Implementing Academic Standards is a groundbreaking document focusing on five important standards-related issues. First, it defines and outlines the critical components of a successful academic standards system using examples of good and bad standards to illustrate key points. In addition, it offers analyses, discussion, and recommendations in four other crucial areas: tests aligned to the standards; categorization of student achievement through performance standards; implementation and accountability systems to guarantee that the standards become a reality in the classroom; and communications strategies to disseminate information about the standards to parents, teachers, and local school officials. Unless each of these components is both present and of high quality, a system of education standards will most likely fail to improve achievement.

Academic content standards set out the essential subject knowledge and skills students must master at defined intervals in their school careers. During the 1990s, a consensus has developed on the qualities that should be embodied in any good set of standards. According to a wide range of experts, a good set of academic content standards, in whatever subject, should be:

1) Rigorous;
2) Intelligible;
3) Measurable;
4) Specific;
5) Comprehensive;
6) Academic;
7) Balanced;
8) Manageable; and,
9) Cumulative.

Examples of academic content standards meeting these requirements include California's math and reading standards and Japan's math standards.

Schools and districts where students fail to meet the standards should be targeted for reforms guaranteed to shake the status quo.

Crafting the content standards is just one part of a comprehensive system. Without an assessment device, there can be no way of knowing if the content standards are being met in the classroom. Further, if the assessment device does not accurately measure the knowledge content, then it will be impossible to determine if the standards are being met. An assessment device tests students'

BEST COPY AVAILABLE

subject knowledge and skills and the results are reported to officials and the public. In deciding upon the type of assessment device to use, policymakers should bear in mind issues such as:

1) Depth vs. breadth;
2) Time and cost of scoring;
3) Ability to generalize;
4) Factual knowledge vs. higher order thinking skills;
5) Memorability;
6) Equivalency; and,
7) Validity.

Performance standards designate the achievement levels on the state test (e.g., "advanced," "proficient," "basic," and "below basic") and what the cut-off scores for the achievement levels will be on the test. It is important that the cut-off scores not be pegged artificially low so that, for example, more students score at the "advanced" level than is warranted by the students' actual knowledge. In crafting performance standards, the following steps should be observed:

1) Set the number of performance standards.
2) Name the performance levels.
3) Provide content and quality of performance at each level.
4) Develop and administer test items.
5) Decide cut scores.
6) Provide student work samples.

A high-stakes implementation and accountability system must be put in place so that local school districts have an incentive to make sure the standards actually make it into the classroom. Such a system cannot consist of more money thrown at districts which underperform. Rather, schools and districts where students fail to meet the standards should be targeted for reforms guaranteed to shake the status quo. Policymakers should therefore consider these implementation and accountability strategies:

1) Performance contracting with outside firms to provide educational services;
2) Merit pay for teachers linked to student achievement on standards-aligned tests;
3) Teacher selection and renewal based on performance;
4) Targeted school-choice demonstration programs in districts where students fail to meet the standards;
5) Improving teacher training programs by increasing content-area requirements; and,
6) Improving professional development for existing teachers by emphasizing standards-aligned content-area knowledge.

Finally, a communications plan must be formulated that informs parents, teachers, local school officials, the media, and the general public about details of the standards, assessments, performance standards, and accountability mechanisms.

PACIFIC RESEARCH INSTITUTE
1979 20 Years of Ideas 1999

This Ideas In Action fact sheet is a digest of a publication entitled: **Developing and Implementing Academic Standards: A Template for Legislative and Policy Reform.** This study is available through PRI's Publications Department for $16.95.

755 Sansome Street, Suite 450, San Francisco, California 94111
Ph: 415-989-0833 Fax: 415-989-2411 E-mail: pripp@pacificresearch.org http://www.pacificresearch.org